THE AMERICAN
MATHEMATICAL MONTHLY

AN OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

DEVOTED TO THE INTERESTS OF COLLEGIATE MATHEMATICS

THE AMERICAN MATHEMATICAL MONTHLY, FOUNDED IN 1894 BY BENJAMIN F. FINKEL, WAS PUBLISHED BY HIM UNTIL 1913. FROM 1913 TO 1916 IT WAS OWNED AND PUBLISHED BY REPRESENTATIVES OF FOURTEEN UNIVERSITIES AND COLLEGES IN THE MIDDLE WEST

VOLUME 89

1982

PUBLISHED BY THE ASSOCIATION
MONTPELIER, VT, AND WASHINGTON, DC

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 1 January 1982

Contents

NOTICE TO AUTHORS

See statement of editorial policy (this issue, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

STATEMENT OF POLICY

Several of the past editors of the MONTHLY published a statement of their policy in their first issue; I continue the tradition. Much of what I want to say was already said four years ago by my immediate predecessor, Ralph Boas, and I cannot say it better; for that reason several passages below are lifted verbatim from his statement.

Since its beginning, the MONTHLY has been dedicated to the advancement and promotion of college mathematics. The founders of the MONTHLY set out to create a publication that would be neither a research journal nor one devoted primarily to educational and pedagogical topics. Following my predecessors, I again endorse these principles.

The important word is *mathematics*. The focus is to be on the reading, the writing, the discovery, the teaching, the problems, the joy, and the people of mathematics. The MONTHLY is expected, in addition, to provide services of more ephemeral value (such as various lists of names, titles of books, and reports of meetings), and it will continue to do so. An easily identifiable and removable Center Section has been especially designed for that purpose.

The traditional departments of the MONTHLY will be continued, in some cases slightly differently, and therefore under slightly different headings. Here follow a few words about each of them.

ARTICLES should be of wide interest, written in an expository manner. All parts of mathematics are welcome, pure and applied, old and new. The level should be accessible to readers who have had no more than a first-year graduate education in mathematics, but, of course, this can only be an average. Some articles are bound to be more advanced and others less, and, as a result, it is to be expected that the average reader will find a few articles too difficult and a few too easy. There is nothing wrong with that; it makes for a magazine far better than one that ruthlessly edits all contributions to an insipid mean value.

The MONTHLY is not the place to publish research papers of interest to specialists only, and it will avoid what is best described as "minor research". This is not intended, however, to exclude the occasional article (or note) containing new results that can be understood by nonspecialists and can be expected to be of wide appeal.

Editors and readers join in asking authors to write correctly and clearly. Beyond the basic principles of good writing, this requires above all that the author show consideration for the reader, who will usually not know as much about the subject as the author does. The prospective reader is especially encouraged if each article has an introduction (from a paragraph to several pages, depending on the nature of the article) saying informally what the article is about and how it fits in with more familiar material. Avoid a tight "definition-theorem-proof" format; the MONTHLY will gladly trade brevity for clarity. Avoid too many definitions, especially in the first paragraph or on the first page, and remember, throughout the article, that most people understand words much more quickly and easily than formulas. Keep your sentences short and simple. Do your best to arrange the article so that it doesn't seem to ramble but hangs together and gives the feeling of an organized structure.

PROGRESS REPORTS are brief reports on more or less recent interesting developments in mathematics. They are not intended to be research advertisements written by their discoverers, but parts of a planned and coherent series, usually written by specially invited contributors.

NOTES continue the department formerly called Mathematical Notes, and, in part, the former Classroom Notes. (Both editors and readers have been having trouble distinguishing between these two departments.) Notes should in general be short papers of one to eight typed pages that give new insights, new or improved proofs of old theorems, brief bits of mathematical folklore that have not found a home in the literature, or (occasionally) new results that are not too technical. The topics should be of wide current interest.

UNSOLVED PROBLEMS should be problems whose statements use only concepts common in

undergraduate mathematics, and for which there is some expectation that progress can be made without terribly advanced methods. The appearance of an unsolved problem in the MONTHLY does *not* mean that its solution will necessarily be a suitable article or note—indeed, the opposite is likely to be true.

TEACHING OF MATHEMATICS can suggest thoughts that are illuminating and inspiring to the amateur, to the student, and to the user of mathematics, as well as to the teacher. This department continues the one formerly called Mathematics Education, and, in part, the former Classroom Notes. It is to contain brief articles and notes concerning the teaching of mathematics at college and introductory graduate level. It is not to contain discussions of those teaching problems and techniques that are common to almost all academic disciplines. The rule of thumb is that if a paper remains meaningful when the word "mathematics" is replaced by "geography", then it is not for the MONTHLY; if, on the other hand, it is concerned with the problem of explaining a specific definition, theorem, or proof, then it is a prime candidate for publication in this department.

PROBLEMS. This department has been an especially attractive and valuable feature of the MONTHLY. It will continue to seek and to publish interesting elementary and advanced problems, and their solutions, in pure and applied mathematics, both classical and modern.

REVIEWS. Through Telegraphic Reviews the MONTHLY will continue to provide as complete a coverage as possible of current textbooks. At the same time extended reviews are planned. They should be mathematical mini-essays suggested by recent additions to the book literature. They are not meant to be book reviews in the archival, bibliographical sense only, but summaries in two or three pages for the non-expert of what an author wrote in two or three hundred pages for the expert. The idea is not just to help the reader decide whether or not to buy the book, but to provide, for the busy specialist in something else, some compromise between knowing nothing at all about the subject and spending a hard month reading the whole book.

P. R. HALMOS, *Editor*

THE DISK WITH THE COLLEGE EDUCATION

HERBERT S. WILF
Department of Mathematics, University of Pennsylvania, Philadelphia PA 19104

The title is somewhat exaggerated, but the calculators-or-no-calculators dilemma that haunts the teaching of elementary school mathematics is heading in the direction of college mathematics, and this article is intended as a distant early-warning signal.

I have in my home a small personal computer. About 500,000 small personal computers have been sold in this country, of which a healthy fraction are owned by individuals. I use mine primarily for word processing (this article was written on it), for writing programs that do various mathematical jobs related to my teaching or to my research, for playing games, for keeping class rolls, etc.

A new program has recently been made available for my little computer, one whose talents seem worthy of comment here because it knows calculus; in fact, as you read these words, some of

The author received his Ph.D. degree in 1958 from Columbia University, taught at the University of Illinois, and has been at the University of Pennsylvania since 1962. His principal research interests have been in analysis: numerical, mathematical, and, in the past several years, combinatorial. He is the author of several books and articles, has been a John Simon Guggenheim fellow, has received a Christian and Mary Lindback award for excellence of undergraduate teaching, and is co-editor-in-chief (with D. E. Knuth) of the *Journal of Algorithms*.

your students may be doing their homework with it.

The program is called muMATH; it was written by the Soft Warehouse, and is distributed in the United States by Microsoft Consumer Products of Bellevue, Washington. It costs about $75 and is supplied on a 5-inch floppy disk with an (inadequate) instruction manual.

The program on the disk does numerical calculation to high precision, or symbolic manipulation of expressions. The numerical calculation, which is less important as far as this article is concerned, is in rational arithmetic and is done with 611-digit accuracy. Thus, for example, when the program is loaded, the question

?30!;

yields the instant answer

@265252859812191058636308480000000

The question

?1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7;

elicits

@363/140

and so forth.

But these are fairly standard calculator-type questions. The first glimmer that a nontrivial intelligence lives on the disk comes with the request for $\sqrt{12}$,

?12 ↑ (1/2);

(the up arrow means "to the power"), whence the response

@2*3 ↑ (1/2)

At least the disk has been to junior high school. The next few samples show it in grades 9–12:

?(X + 2*Y) ↑ 3;

@12*X*Y ↑ 2 + 6*X ↑ 2*Y + X ↑ 3 + 8*Y ↑ 3

?COS(5*Y);

@ −20*COS(Y) ↑ 3 + 16*COS(Y) ↑ 5 + 5*COS(Y)

(Tschebycheff polynomials, anyone?).

The disk, however, has graduated from high school. Here it is in a freshman calculus course. To differentiate $x \sin x$ with respect to x just ask

?DIF(X*SIN(X),X);

to obtain

@X*COS(X) + SIN(X)

At the risk of some eyestrain, we might even ask it to

?DIF((X ↑ 3 + COS(X)) ↑ (1/2),X);

to which it replies

(3*X ↑ 2*(X ↑ 3 + COS(X)) ↑ (1/2)/2 − (X ↑ 3 + COS(X) ↑ (1/2)*SIN(X)/2)/ (X ↑ 3 + COS(X))

In fairness to it, I should say that there are some control variables that tell it how we would like our expressions to be simplified, and I didn't experiment to find the best settings of these variables; so there probably is a way to get the answer in more elegant form.

Since we can differentiate with respect to x, it seems that the thing should know how to differentiate partially; and so we ask

```
?DIF((1 + T*X ↑ 2) ↑ (1/2),T);
```

Our burgeoning inferiority complex is promptly reinforced with

```
@X ↑ 2/ (2*(1 + X ↑ 2*T) ↑ (1/2))
```

Next we come to integration, and the picture is a bit uneven. To integrate (antidifferentiate) $1/(x + 7)$ with respect to x, enter

```
?INT(1/ (X + 7),X);
```

and of course,

```
@LN(7 + X)
```

Our self-esteem recovers somewhat, though, when we note that the simple question

```
?INT(X*SIN(X),X)
```

gets no helpful response (it can't do everything) but plunges again when the answer to

```
?INT((1/ (3 + 4*X + 5*X ↑ 2)),X):
```

appears instantly as

```
@ATAN(5*X/11 ↑ (1/2) + 2/11 ↑ (1/2))/11 ↑ (1/2)
```

which is its way of saying

$$\frac{1}{\sqrt{11}} \arctan\left(\frac{5x+2}{\sqrt{11}}\right).$$

There's yet another dimension to this disk. So far we've discussed its calculator mode of operation, in which we ask just a single question and it gives just a single response, however clever. But a whole computer language called muSIMP (similar to LISP) also lives on the disk, and we can write programs in muSIMP that use all of the capabilities above plus decision-making, looping, etc. The possibilities are numerous, and here are a few of them.

Suppose we want to develop a function, given as an expression, in a Taylor series about 0. We could evaluate the expression at 0, then replace the expression by its derivative (!!) and loop back. A complete subroutine to print out the successive derivatives of a given function (EXPR) evaluated at 0 might look like this:

```
?FUNCTION SERIES (EXPR),
LOOP
L:EVSUB(EXPR,X,0),
PRINT(L),
SPACES(2),
X:'X,
EXPR:DIF(EXPR,X),
ENDLOOP,
ENDFUN$
```

Inside the LOOP, we first evaluate EXPR by replacing each X by 0 and call the result L. Then

we PRINT L and two spaces. The next instruction is a technicality that we will not discuss here, and finally we replace EXPR by its derivative.

To use this, we call for

?SERIES(SIN(X)):

and get

@0 1 0 −1 0 1 0 −1

etc. More interesting is to see the sequence of Bell numbers $b(n)$ $(n = 0, 1, 2, \dots)$, where $b(n)$ is the number of partitions of a set of n elements. The exponential generating function is well known to be $\exp(\exp(x) - 1)$, and so we ask

?SERIES(#E ↑ (#E ↑ X − 1));

and, as we watch, the screen gradually fills up with

@1 1 2 5 15 52 203 877 4140 21147

(after 4 minutes). By stopping the program here, we can peek at the symbolic ninth derivative of $\exp(\exp(x) - 1)$, and so, fearing the worst, we type

? EXPR ;

Instead of reproducing the output exactly as it appeared, I'll render it into standard notation as

$$\exp(e^x - 1 + x) + 255\exp(e^x - 1 + 2x) + 3025\exp(e^x - 1 + 3x)$$
$$+7770\exp(e^x - 1 + 4x) + \cdots + 36\exp(e^x - 1 + 8x) + \exp(e^x - 1 + 9x).$$

The coefficients seem to be Stirling numbers $S(9, k)$ of the second kind, and so we are led to conjecture that the nth derivative of $\exp(\exp(x) - 1)$ is

$$\exp(\exp(x) - 1) \sum_{k=1}^{n} S(n, k) e^{kx}$$

and a proof follows quickly by induction.

For a parting volley, note that we can ask for the Taylor series in powers of x of a function that depends on x and t, say, and the output coefficients will then be functions of t, printed symbolically. Thus, to see the Legendre polynomials (times $n!$), we inquire

?SERIES((1 − 2*X*T + X ↑ 2) ↑ (−1/2));

and across the screen there go marching

@1 T −1 + 3*T ↑ 2 −9*T + 15*T ↑ 3 9 − 90*T ↑ 2 + 105*T ↑ 4

Programs that do symbolic manipulation of mathematical expressions are not new. The MACSYMA program of MIT has been doing it for years. What is new is the sudden mass availability of a program with these capabilities, and the promise that more of the same is in prospect.

This year for the first time there are widely available pocket computers: objects about the size of calculators that have thousands of bytes of memory and speak BASIC. These pocket computers are not yet quite powerful enough to handle a sophisticated program like muMATH. In a few years, though, they probably will be.

As teachers of mathematics, our responses might range all the way from a declaration that "no computers are allowed in exams or to help with homework" to the if-you-can't-lick-'em-then-join-'em approach (teach the students how to use their clever little computers).

Will we allow students to bring them into exams? Use them to do homework? How will the content of calculus courses be affected? Will we take the advice that we have been dispensing to

teachers in the primary grades: that they should teach more of concepts and less of mechanics? What happens when $29.95 pocket computers can do all of the above and solve standard forms of differential equations, do multiple integrals, vector analysis, and what-have-you?

Excuse me if I don't have answers. I wanted only to raise the questions and beat a hasty retreat.

A DOWN-TO-EARTH VIEW OF MATHEMATICS

F. F. BONSALL
Department of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, Scotland

In his recent article "Applied Mathematics Is Bad Mathematics," Halmos [**3**] dangles an attractive bait in the fishpond, maintaining that applied mathematics is different from pure mathematics and that much of it is bad mathematics. I do not propose to nibble that bait but to tweak his line with a related question. How do we know that a piece of mathematics is good or bad? For Halmos, mathematics is an art and merit is a matter of aesthetic judgment. This I find an unsatisfying answer. It fails to explain the near unanimity in judgment of mathematical merit, in striking contrast to judgments of other works of art. Moreover, this unanimity is not just a passing fashion, as so often with other arts, but seems to endure. Also mathematics differs from other arts in that there are necessary conditions that must be satisfied by a piece of good mathematics—for example, it must be correct and nontrivial. Are there any such criteria for paintings or poems?

I do not seek to show that the merit of a mathematical work can be measured in terms of its application outside mathematics; though such application may well be part or even the whole of its merit. The fact that mathematics is necessary for many sciences and for engineering is, in global and historical terms, a vital part of the justification for our work and an important source of motivation; but, when we look at an individual theorem, applicability is seldom a relevant consideration. For example, mathematicians surely agree that the prime number theorem (the number of primes not exceeding n is asymptotic to $n/\log n$) is in the highest class of mathematics; but is it applicable? I aim to show that nevertheless it is possible to take a down-to-earth view of mathematics in which merit does correspond to utility of a certain kind, namely, its contribution to a science of mathematics. I must warn that everything in this article will be very simple-minded and devoid of linguistic subtlety—I am no sort of philosopher but merely a mundane mathematician.

A Science of Mathematics. Mathematics is an activity of human beings. I am not saying that it is impossible for other animals to be mathematicians; whales, for example, might have the necessary brain power. My point is only that mathematics is an activity and is subject to the usual human limitations and defects. It is mainly an activity of individuals and there seems no place for the large-scale teamwork to be found in some experimental sciences. It is also true that we owe a large part of the progress of mathematics to a very small number of outstandingly original

Since 1947, the author has held appointments at the University of Newcastle upon Tyne and at the University of Edinburgh, where he is currently a professor, and visiting positions at Oklahoma State University, The Tata Institute for Fundamental Research, and Yale University. He was awarded the Berwick Prize of the London Mathematical Society in 1966 and was elected F.R.S. in 1970. His research work has been in analysis, mainly in the theory and application of Banach algebras, and he is the joint author of three books.—*Editors*

mathematicians who have taken great strides into the unknown. But mathematics is also a collaborative undertaking, in that even the greatest mathematician needs to use the work of others.

All human beings are fallible, and so errors inevitably occur, ranging from serious errors of understanding to minor misprints, which are very difficult to eliminate because of the tedium of proofreading. In ordinary prose there is enough redundancy for a scattering of misprints to be no more than a minor nuisance; but there is little redundancy in a mathematical formula and a single misprint may well change the meaning completely.

It might be interesting to attempt to classify the errors of professional mathematicians. They range from trivial misprints to thumping great mathematical howlers. A not uncommon sort of error involves some obvious foolishness like writing $2 < 1$. In such a case the author may have had in mind $1 < 2$, in which case the reader can make the correction without difficulty. But after writing $2 < 1$, the author may have gone on to base the subsequent argument on this powerful inequality. Then the reader will have trouble. A more serious sort of error is one which invalidates some important conclusion but does not stem from some obvious foolishness but rather from some subtle misunderstanding. The fact that errors exist in almost all major publications would seem to make it impossible for mathematicians to use the work of others without the entire subject collapsing into a morass of uncertainty. And so it would, had not mathematicians devised a scientific method through which the important theorems become known with a degree of certainty that is lacking in all other human activities.

It is well known that experimental and observational scientists have devised a scientific method whereby they can use the work of others. In this method, results and observations are published together with sufficient information about the experimental techniques to enable other experts in the field to verify the results for themselves. In this way, errors, and even fraudulent claims, are eliminated. It is sometimes argued that this picture of the scientific method is illusory in that the great majority of experiments are not repeated. It is no doubt true that many results are of a routine nature, appear entirely unsurprising to other experts in the field, and may not be tested by repetition of the experiment. But the important results, those that conflict with the expectations of experts or change the subject radically, will be tested.

Mathematicians have also devised an effective scientific method appropriate to their subject; but, perhaps because the word *science* is usually attached to the experimental and observational sciences, this is less widely understood. This scientific method for mathematics involves the publication of results in the axioms-theorem-proof form. Introduced by the classical Greeks more than 2,000 years ago, this form is now almost universally in use. The explicit statement of the axioms (or definitions, I make no distinction) enables another mathematician to decide whether the theorem is applicable to his own problem, and the proof enables him to check that the theorem is correct. The proof also provides some very useful redundancy. Mathematics has to be written in ordinary language supplemented by formulae, and it is often written with insufficient care or with "abuse of notation." Every mathematician is familiar with the experience of needing to dip into the proof before being able to understand what is being claimed in a theorem.

By the *science of mathematics*, I mean the collaborative activity of mathematicians publishing their work in this form, each author accepting full responsibility for the correctness of the whole of his publication *including the results that he quotes from other authors*. Pursued in this way, mathematics can remain permanently healthy, significant error being eliminated automatically. Everything in a publication must be based on the individual understanding of the author, nothing being accepted on authority, no matter how distinguished. Since the author is staking his reputation on the work of other authors that he uses, that work is checked by somebody who has the strongest incentive to detect the errors. This is far more effective than relying on referees, reviewers, and other more or less passive readers. A second mathematician using the work of another may himself fall into error in following the proof, but at least he has a strong emotional

drive to avoid such error. In fact, if human nature is what it is commonly supposed to be, he may well have a stronger motive to detect error in that work than the original author had. It is true that this scientific method will not detect the errors in theorems that are never used. But that is unimportant; such a theorem is a dead branch anyway.

Although I have described the science of mathematics as a collaborative activity, please do not misunderstand me; it is not an activity of society. Laws, judges, priests, politicians, prisons, wars, votes, strikes, mob violence, hostage taking, all equally fail to affect the validity of a theorem.

The nature of the axiomatic method has frequently been misunderstood even by distinguished mathematicians. It has sometimes been interpreted as an attempt to start from some primitive axioms of set theory and then to build the whole edifice of mathematics on this foundation by rigorous logic. This is the very opposite of the axiomatic method as understood by the real live mathematician. Instead of tying him down to some dubious foundations the axiomatic method allows him to dance in the air by taking anything he pleases as his starting point.

The words and symbols used in mathematics must be defined, but can only be defined in terms of other words and symbols, and so in the last resort cannot be defined at all. Thus when Jones uses Smith's theorem he may not know precisely what Smith meant. But this is of no importance; if Smith's theorem is correct for Jones's concepts, Jones can go ahead and use it. Likewise we need not concern ourselves with the logical language or rules of proof that are used. It may be that Jones, when he reads Smith's proof, will sometimes decide that the proof does not satisfy his own logical requirements. In that case his remedy is either to find a new proof or to reject Smith's theorem.

I hope that I have made it clear that in speaking of the science of mathematics I do not envisage some vast, mutually compatible assembly of rigorously established mathematical facts—like some gigantic Bourbaki. What I mean is a certain attitude of mind and mode of operation among working mathematicians. My emphasis on the role of the living mathematician might lead to the misunderstanding that I accept idealistic philosophy. Not a bit of it; I am not a philosopher and agree with G. H. Hardy [**5**]. "Pure mathematics, on the other hand, seems to me a rock on which all idealism founders: 317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but *because it is so*, because mathematical reality is built that way."

The Nature of Proof. I am not concerned with logical refinements but only with the practical necessities of the science of mathematics as practiced in the real world. In using somebody else's theorem, I cannot rely on the distinction of the author or the prestige of the journal but must check its correctness for myself. I must myself understand how the theorem follows from the axioms with the aid of such other theorems as I have already checked. This act of understanding cannot be performed for me by anyone else, still less by a computer, and I have only one lifetime in which to work. Even if I am fortunate enough to be free from other work and distractions, the greatest enemy to accuracy is boredom. Thus to perform its health-giving function a proof must be understandable by a real live mathematician in a reasonably short time. What is reasonable will depend on the importance of the theorem; we will tolerate more in a proof of the Riemann hypothesis than in a proof of a run of the mill result.

How we check correctness is up to us; it is our own reputation that is at stake. Some have the patience to do a meticulous line-by-line verification, others will rely mainly on checking with examples and on their experience and intuition to direct them to check the crucial steps. The meticulous detailed check is not necessarily superior. Not only is boredom a deadly enemy, but some major gap may go undetected. Everything that is written down may be perfectly correct but it may still leave the theorem unproved. The best check involves both methods.

Our real live mathematician has only a limited knowledge of mathematics, so a proof should use economy of force. It should not invoke deep results if a little elementary calculus will do the trick. But on the other hand there is no merit in an elementary argument if it becomes long and

boring. Our real live mathematician would prefer to master some difficult tool than to endure prolonged tedium. The Erdős-Selberg elementary proof of the prime number theorem was a most remarkable tour de force, but no real live mathematician would use it in preference to the function theoretic proofs.

Live mathematics is that body of mathematical theorems that is currently understood by living mathematicians. A substantial trace of this mathematics is left behind in a fossilized form in publications, just as the coral reef is left by the polyps. Standards of rigour and the active interests of mathematicians change with time, and so in practice it is only quite recent publications that are actively used. The proportion of the human race that understands the notion of mathematical proof is quite small and the notion of proof does not seem to come naturally to children. In fact there is much greater readiness to accept authoritative statement than to undertake the effort needed for understanding. It is not at all hard to envisage the decline of our civilization into an authoritarian state in which mathematical understanding disappears altogether and mathematicians are replaced by priestly persons interpreting these mysterious writings. Indeed, perhaps this happened in some ancient civilizations. It is not even necessary to require the suppression of mathematical understanding by authoritarian force. Perhaps the mathematical powers of the human race could be atrophied by soft and easy living and reliance on calculators at school.

It would be splendid if all proofs could be intelligible to all professional mathematicians. In our real world, understanding a proof needs a great deal of prior knowledge, at least in the more highly developed branches, though we can fairly demand that this prior knowledge be available from published sources. Our concept of the science of mathematics requires that we should personally understand the proof of a theorem before using it. The need to master a great deal of difficult mathematics before we can use a theorem severely limits the speed at which we can progress.

My description of the science of mathematics may be fully applicable only to pure mathematics. Many applied mathematicians naturally feel less responsibility for the correctness of the theorems from pure mathematics that they use. Also it seems that mathematics may be splitting into subsciences. For example, I am told that number theorists find it necessary to use results from algebraic geometry that they are unable to check for themselves. Because of the brevity of human life they have to content themselves with reading enough of the proofs to be sure that they are interpreting the results correctly. If this were to become a widespread practice it would be fatal for mathematics. So long as it is more or less isolated it may not matter over much, fallacious results will usually be dead branches or will eventually lead to obvious absurdities.

From our mundane standpoint, expository writing can be seen to be a necessary part of our science. It is unlikely that the simplest proof of a theorem will be found at first and simpler proofs are obviously valuable. Even a new proof that is not technically simpler but bases the proof on some other background of knowledge can be useful in making the theorem more widely available. To be useful, expository works should be written in reasonably short, self-contained units. The authoritative encyclopaedia has no place in our science. The more authoritative, the worse, because the more likely to tempt us to accept external authority in place of our own understanding. The real live mathematician never reads a long book from cover to cover but uses books to help him understand some specific theorem or group of theorems. Nothing is more frustrating than trying to use some vast work in which the notation is magnificently consistent but in which nothing can be understood without what seems to be an unending regression.

Good and Bad Mathematics. Is this just a matter of aesthetic taste? I have already said that I think not and that we can relate merit in mathematics to the practical needs of our science. In the first place, we all ask of good mathematics that it should make some nontrivial advance either in discovering new theorems or in giving us new understanding of old ones. If a mathematical theorem is to avoid becoming a dead branch, it must stimulate the interest of real live

mathematicians. This can be done either by contributing to some branch which already interests other mathematicians or by developing a new branch with results so striking or surprising that they attract attention. A great part of the merit of good mathematics is related to this need to stimulate interest. For instance, we like to be surprised and are therefore attracted by theorems that look almost paradoxical. I have remarked already that all mathematicians agree that the prime number theorem is in the highest class. It is worth while considering why this is.

Let $\pi(n)$ denote the number of primes not exceeding n; then the prime number theorem tells us that

$$\lim_{n\to\infty} \frac{\pi(n)\log n}{n} = 1.$$

The following empirical data can be found in Hardy and Wright [**6**].

n	10^3	10^6	10^9
$\pi(n)$	168	78,498	50,847,478
$n/\log n$	145...	72,382...	48,254,942...
$\pi(n)\log n/n$	1.159...	1.084...	1.053...

I need hardly say that I have not verified the entries in this table, but I am only writing about mathematics, not doing it. These empirical facts are of course of no help toward a proof of the theorem, but they do have the merit of indicating the hopeless amount of calculation involved in trying to obtain convincing empirical evidence for the truth of the theorem, even with the most powerful computers.

Anyone with some limited experience of prime numbers will know that they seem to be distributed in a very irregular fashion, sometimes close together and sometimes with large gaps. Yet the prime number theorem asserts a simple regularity in the limit. Thus the theorem has a stimulating unexpectedness. Another merit of the theorem is the extreme simplicity of the statement, which only involves the most elementary notions. There are however many theorems that are both surprising and simple to state. What makes the prime number theorem one of the great theorems is the extreme difficulty of proving it.

The prime number theorem was conjectured by Gauss in 1793, but the story of the proof goes back to about 1740 when Euler discovered a formula for the zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ in terms of the prime numbers. Around 1850 Tchebychef proved that there exist positive constants A, B such that

$$A\frac{n}{\log n} \leqslant \pi(n) \leqslant B\frac{n}{\log n}.$$

In 1859 Riemann showed that the key to the study of $\pi(n)$ is the location of the complex zeros of the zeta function, but it was not until 1896 that Hadamard and de la Vallée Poussin proved independently that none of these complex zeros have their real parts equal to 1, and thus completed the proof of the theorem. An interesting short account of the history, from which these remarks have been extracted, will be found in Hardy [**4**]. For a full account of the prime number theorem including several of its proofs, see Ayoub [**1**].

In what sense is the prime number theorem important? The long search for the first proof and, even after that, the search for simpler proofs must have provided a most valuable stimulus to the development of analytic number theory and related parts of analysis. Even today most mathematicians would be happy to find a substantially simpler proof. In another sense the theorem is probably not important. I have already mentioned that it does not seem to be useful for applied mathematicians. But I also suspect that it is seldom used by pure mathematicians. It is typical of many difficult theorems in being important only as a goal and inspiration. This indicates that

mathematical importance may be of several kinds. The prime number theorem should be contrasted with (say) Cauchy's theorem and the Hahn-Banach theorem, which are quite superficial and easy to prove but are as useful to an analyst as a spade to a gardener. This second kind of theorem is obviously essential for the science of mathematics, but we should not decry the other kind. The real live mathematician must have his inspiration; but for the existence of such brilliant achievements he would not dare to attack the really difficult problems. It tends to be the case that the really useful theorems for applications in both pure and applied mathematics are the elementary theorems, but it is not true that hard theorems always lack application.

There is an interesting discussion of merit in mathematics by an outstandingly creative mathematician in Hardy's essay "A Mathematician's Apology" [**5**]. This was written in 1940 when Hardy was much depressed by the state of the world and perhaps also by the decline of his own creative powers; and as a result it stresses the uselessness of real mathematics, an emphasis that may have injured the reputation of pure mathematics. To Hardy, real, interesting mathematics is useless, useful mathematics is trivial or dull—Hogben stuff. In the main he finds the real mathematics within pure mathematics, but it is instructive to note that he includes relativity and quantum mechanics in the real mathematics and counts Maxwell, Einstein, Eddington, and Dirac among the real mathematicians.

Hardy seems to take an uncompromising stand that merit in mathematics is determined entirely by its aesthetic appeal to mathematicians. And yet when he examines the difference between mathematics and a chess problem we have: "The best mathematics is *serious* as well as beautiful." After stating and proving two elementary theorems, he draws attention to their unexpectedness, inevitability, and economy. These are indeed merits, but are they purely aesthetic? If "aesthetic" is to be defined so as to cover everything that stirs a mathematician, then I surrender. But in figure skating the judges are expected to draw a distinction between technical and artistic merit.

I do not belong to the school that finds nearly all new mathematics bad. On the contrary, I find the general level of mathematical publication to be at least as high as ever before. It cannot be denied that there are some thin and pointless generalizations and some cottage industries. But we should not worry too much about this; it is better for a mathematician to stay alive by attempting something than to give up the attempt. At the other extreme, and less widely recognized as bad, is bad mathematics done in the name of solving famous problems. It is natural and desirable that mathematicians should be attracted by famous unsolved problems and should make great efforts to solve them. Such efforts bring new and valuable methods into mathematics, but it is well to keep some sense of proportion. The problem itself may have very little intrinsic importance. If such a problem, the four color problem, for example, is solved by some clever new idea, that is magnificent; but a solution by a cumulative application of existing methods may do nothing more than demonstrate the cleverness of the solver. It is worse still if the solution involves computer verification of special cases, and in my view such a solution does not belong to mathematical science at all.

It is no better to accept without verification the word of a computer than the word of another mathematician. In fact the tedium of routine tasks makes programming errors exceedingly probable. We cannot possibly achieve what I regard as the essential element of a proof— our own personal understanding—if part of the argument is hidden away in a box. Very recently it has been claimed that the classification of finite simple groups has been completed with some major assistance from computers. I do not know enough to judge how essential the role of the computer has been, though I am told that a major part of the problem for which computers were being employed was in fact solved mathematically. Perhaps no great harm is done, except for the waste of resources that would be better spent on live mathematicians, so long as we do not allow ourselves to be content with such quasi-proofs. We should regard them as merely a challenge to find a proper proof.

Perhaps we are seeing the birth of a new kind of computer-assisted quasi-mathematics, but it has no place in the science of mathematics and if it is to survive must develop its own scientific method—perhaps more akin to the experimental sciences.

When we see bad mathematics done for the sake of solving famous problems we are seeing the same sort of phenomenon that Halmos observes in applied mathematics. However, just as applied mathematicians need the stimulus of applicability to the physical world, so many other mathematicians need the spur of fame. Since we are all human, it may be that lacking such incentives we might achieve miserably little.

All mathematicians are agreed that problems are the life-blood of mathematics. But their value lies in stimulation of good work. Solution by bad methods robs the world of a good problem. To some extent it eases the search for a good solution, but at the same time it removes much of the excitement.

Then there is jumping on band wagons. There is usually not much harm in it, and the excitement of the chase helps to provide the necessary stimulus. But it is unhealthy for mathematical science when we exploit some immensely difficult theorem (the Feit-Thompson theorem [**2**], for example) long before we can verify its correctness. In the case of such a spectacular theorem we can perhaps be reasonably confident that, if it were eventually found to be false, that falsity would become so well known that the resulting errors would eventually be eliminated. But a willingness to act in this way in the interest of speed weakens our resolve and may eventually be fatal—like too much riding in cars when we would be healthier on foot or at least on bicycles. Other more obvious kinds of bad mathematics, slipshod inaccurate work and too ready an appeal to "it is easy to see" are also due to too much haste. I must clarify my point here. My objection to excessive haste applies only to hasty publication. The active mathematician learning of the existence of some exciting new theorem impinging on his field of competence will wish to explore its consequences at once, and there is not the least objection to his doing so with all possible speed. A great deal has to be done, the creative part of a mathematician's work, before the stage of publication comes and the brakes have to be applied.

I am not asking for less research but for more. If we find it difficult to understand some result that we want to use, we should not run away from it but should persevere until we really understand it. There is quite a good chance that in so doing we shall find some underlying simplicity that has been hidden under complications. If so, we shall have made a double contribution to our science; whereas if we go ahead and publish without such understanding we risk making a negative contribution. Again, our difficulty may stem from an error in the result; and if that error is nontrivial its discovery may lead to some very interesting new mathematics.

My description of the science of mathematics is by no means intended to describe the whole of mathematics. It describes the process by which professional mathematicians seek to eliminate the serious errors. It does not attempt to describe the creative act. Moreover it must be remembered that the discoverers of some of the most brilliantly original mathematical ideas have not been professional mathematicians. Such discoveries may well lack rigorous proof at first and then there is important work to be done by professional mathematicians in bringing them within our science. For a wider view I cannot do better than refer to the articles "Mathematical Creation" by Henri Poincaré and "The Mathematician" by John von Neumann, both of which can be found conveniently in Newman [**7**].

I believe that mathematics is one of the most important of human activities. It is not just a game, though we enjoy playing it. It is not just an art, though at times it is superlative art. It is not the construction of chains of silly little deductions as philosophers seem to imagine. Mathematics is an enduring down-to-earth science concerned with the discovery of amazingly deep and unexpected theorems. It is not easy to convince our fellow citizens of this and so we cannot expect a lavish supply of public funds for our support. So let us avoid wasting those funds on pseudo mathematics with computers and use them instead to support a few real live mathematicians.

References

1. R. Ayoub, An introduction to the analytic theory of numbers, Amer. Math. Soc. Surveys, 10 (1963).
2. W. Feit and J. G. Thompson, Solvability of groups of odd order, Pacific J. Math, 13 (1963) 775–1029.
3. P. R. Halmos, Applied mathematics is bad mathematics, in Mathematics Tomorrow, L. A. Steen, ed., Springer, New York, 1981, pp. 9–20.
4. G. H. Hardy, Prime numbers, British Association Report, 10 (1915) 350–354.
5. ______, A Mathematician's Apology (with a Foreword by C. P. Snow), Cambridge University Press, 1977.
6. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 5th ed., Oxford University Press, 1979.
7. J. R. Newman, ed., The World of Mathematics, vol. 4, Simon and Schuster, New York, 1956.

COMPOUND PERFECT SQUARES

A. J. W. DUIJVESTIJN
Twente University of Technology, Enschede, Netherlands

P. J. FEDERICO
3634 Jocelyn Street, N.W., Washington, DC 20015

P. LEEUW
Twente University of Technology, Enschede, Netherlands

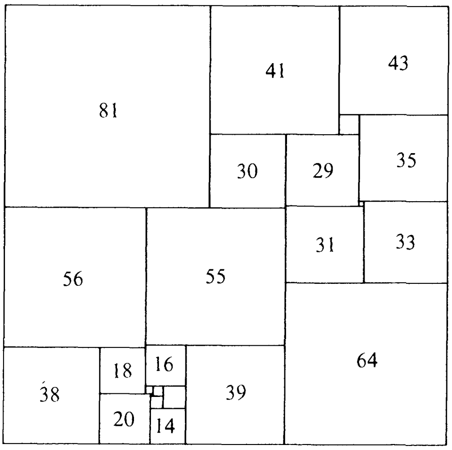
1. Introduction. Perfect squared rectangles came before perfect squared squares. A *squared rectangle* is a rectangle dissected into squares, the *elements* of the dissection. The number of elements is the *order*. If no two elements are equal in size, the dissection is *perfect*, a *perfect rectangle*; otherwise it is *imperfect*. The dissection is *simple* if there is no subset of elements arranged in the form of a rectangle; otherwise it is *compound*. In the latter case the squared rectangle contains a *subrectangle*.

The same terminology applies to squared squares, which may be perfect or imperfect and simple or compound. Since compound perfect squares are subject to an additional condition, it would seem that they would be more difficult to obtain than simple perfect squares and that they would be fewer in number for any given order in which either exist. But while this last statement may be borne out by the few comparative data that exist, it turns out that it is the very restriction on compound squares that enables them to be constructed much more readily than simple squares.

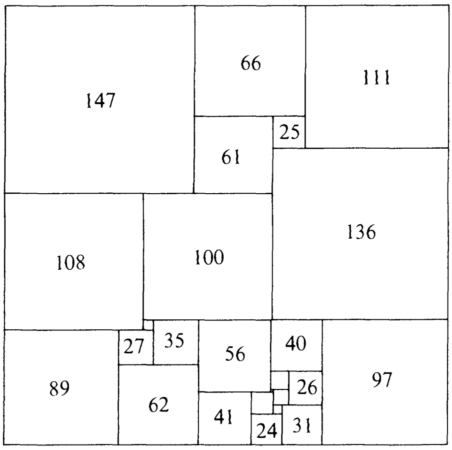
A. J. W. Duijvestijn is Professor in Computing Sciences at Twente University of Technology. He received his Ph.D. from Eindhoven University of Technology under C. J. Bouwkamp; the dissertation was on the subject "Electronic Computation of Squared Rectangles." He also holds a Master's degree in Electrical Technology and in Mathematics from the University of Amsterdam.

P. J. Federico is a member of the bar of the District of Columbia and of the Supreme Court of the United States. From 1923 to 1970 he was employed by the U.S. Patent Office; from 1950 to 1972 he was a Professorial Lecturer in Law at the National Law Center of George Washington University. His mathematical interests are in graph theory and polyhedra, in which fields he has published several papers, and in the history of mathematics.

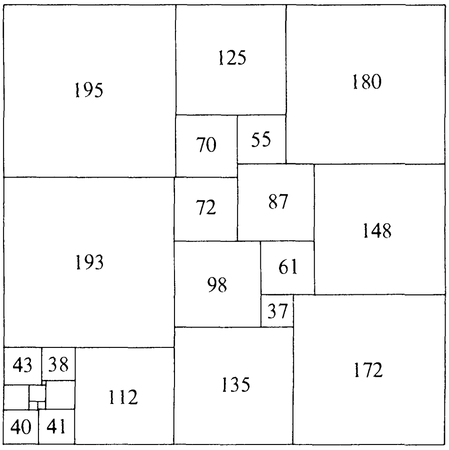
P. Leeuw received his Bachelor's degree in Mathematics from Twente University of Technology in 1979. He is now with Pandata B. V. in Rijswijk, the Netherlands.



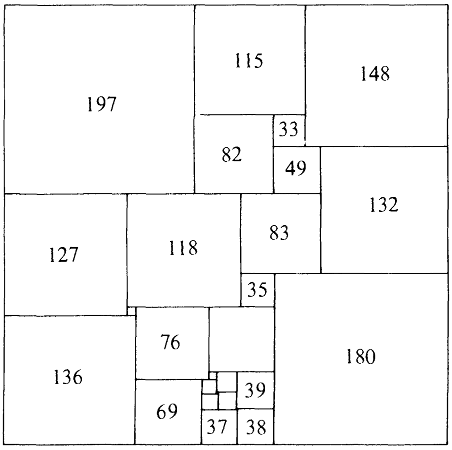
(a) 24-175 [12]



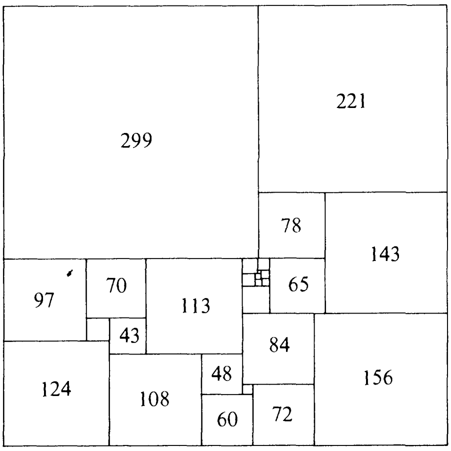
(b) 25-344 [8]



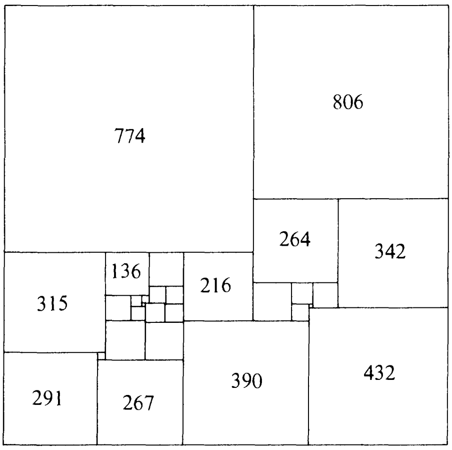
(c) 26-500 [P.J.F.]



(d) 27-460 [18]



(e) 28-520 [19]



(f) 30-1380 [P.J.F.]

FIG. 1. Some known compound perfect squares

The construction of simple perfect squares is still intractable to any direct general theoretical treatment. There are a few special methods that have produced simple squares of particular types (some of order 25, 26, 27 and a few of order 31 and higher), but the only general method is still basically to construct squared rectangles up to a given order and hope that some will turn out to have equal sides. It was not until 1978 that Duijvestijn [**7**] succeeded in finding the simple perfect square of lowest order, 21 squares, using computer methods. The lowest-order compound perfect square known, 24 squares, was discovered in 1948 by Willcocks [**12**], and numerous ones of higher order have been constructed. Methods of producing compound perfect squares are reviewed in [**8**], which lists all those of low order (up to 28) known at that time. A review of the entire subject is in [**9**].

The objectives of the work described in this paper were to establish the lower limit of the order of compound perfect squares and to find all that can exist up to and including order 24. The main results are:

1. There are no compound perfect squares below order 24.
2. There is one and only one compound perfect square of order 24.

Incidental results for higher orders are described in Sections 5 and 6. The method used will be described in some detail in an expository manner, with some description of compound squares.

Compound squares are considered separately in two types: first, those that have only one subrectangle; and, second, those that have two subrectangles not having any element in common. In either case the subrectangle or subrectangles can be compound. It is the maximum subset of elements arranged in a rectangle that is the subrectangle. A square with a subrectangle that itself is composed of or contains two sub-subrectangles not having any element in common is here placed with the first type. Perfect squares with three disjoint subsets of elements arranged in three subrectangles are not considered, since any such must have at least 27 elements, the smallest possible number of elements for a perfect squared rectangle being 9 (the lowest-order one known has 38 elements).

2. Type 1 Compound Perfect Squares. Fig. 1 illustrates a variety of known compound perfect squares of the first type. The length of the sides of the elements of a squaring are conventionally given as integers without any common factors; these are indicated in the center of the square element when possible (the lengths not given can be determined readily). The order of the square and the length of its side are indicated under each. Prior to the present work, 133 squares of this type up to order 30 were known; order 24, 1; 25, 2; 26, 10; 27, 19; 28, 33; 29, 49; 30, 19; and some others of higher orders. Most commonly, the subrectangle is located in a corner of the square, as in (a), (b), (c), (f); this is the case with 124 of the 133. It may be imprisoned in a side of the square, as in (d), 8 of the 133, or it may be wholly inside the square, as in (e). The subrectangle may be a simple perfect rectangle, as in (a), (d); these account for 101 of the 133. It may be a trivially compound rectangle, as in (b), (c), (e), or a nontrivial compound, as in (f); this is the case with 2 of the 133 (see Section 4). The subrectangle in Fig. 1(a) can be inserted in four different ways; the four squared squares thus produced are not considered essentially different and are counted as only one perfect square. With rotations and reflections this square can be represented in 32 different ways. If the subrectangle itself is compound, there are a greater number of ways; thus the squares of 1(c) and 1(e) can be represented in 128 different ways. A standard order of giving the elements of a square and the subrectangle has been adopted in the listing of results, and the squares have been drawn by computer in accordance with this order.

The method used here for constructing squares of this type is essentially that introduced in [**8**]. A square is dissected into unequal squares and one rectangle. (The rectangular element is not to have an entire side in common with the entire side of an adjacent square element; for, if it did, the two would be merged into one rectangular element.) Then a squared rectangle, simple or compound, having the same ratio of sides as the rectangular space is sought, to be fitted into this space. The square divided into squares and one rectangle is referred to as a D (deficient squared

square) and the number of squares as its order; thus D_6 refers to a square divided into 6 unequal squares and one rectangle.

Six is the lowest possible order for a D with no equal elements; hence perfect squared rectangles up to order 18 are needed for a complete canvass up to and including order 24 squared squares. Nine is the lowest possible order for a perfect squared rectangle, and hence the D's up to D_{15} are also needed for a complete canvass to order 24.

The procedure consists of three parts: (1) establishing a list of the essential D's, from D_6 to D_{15}; (2) establishing a list of the essential perfect squared rectangles, simple and compound, from order 9 to order 18; and (3) comparison of the ratio of sides of the rectangular spaces of the D's in (1) with the ratio of sides of the rectangles in (2). For the comparison, each list is sorted according to the nondecreasing ratio of the short side to the long side of the rectangular space of the D's and the ratio of the short side to the long side of the rectangles, respectively, and the matching pairs are readily found. Each matching pair gives a compound squared square that must then be checked for duplicate elements to determine if it is perfect.

Utilization of the two lists, in the two ranges mentioned, results in the complete canvass of all possibilities up to and including order 24 squares: order 25 is almost, but not quite, completely covered (2 out of 11 possible combinations not being represented), and the incompleteness increases up to order 33, which is the most incomplete (only one of 19 possible combinations being represented). Since the combinations that could result only in squares of order 25 and higher are not complete in any event, some of these combinations were not completely canvassed, as will be pointed out later.

3. Generating the D's. The basic paper of Brooks, Smith, Stone, and Tutte [**5**] gives the theory of producing simple squared rectangles from 3-connected planar graphs, called *c*-nets in that and other papers, by means of an electrical analogy. They point out that the method can be extended to produce rectangles divided into rectangles, and many of the propositions in the paper are not limited to square elements. The electrical analogy as applied to a dissection that may include rectangular elements is described by the following illustration.

In Fig. 2, (a) shows a rectangle dissected into 7 rectangles (here the word rectangle may include those with equal sides). Note that no two of the rectangular elements have an entire side in common; more broadly stated, the rectangle does not contain any subset of elements, more than one and less than all, arranged in a rectangle. It is referred to here as a simple rectangled rectangle. The rectangular elements have been numbered (1) to (7) and the horizontal lines numbered 1 (the top side) to 5 (the bottom side). Let H be the length of the horizontal sides of the whole rectangle and h_i $(i = 1$ to $7)$ that of the rectangular elements, and V the vertical sides of the whole and v_i of the elements.

The configuration of Fig. 2(a) is represented in 2(b) by a two-terminal electrical network (the dotted line is to be ignored for the moment). The top and bottom lines, 1 and 5, of (a) are represented by the poles 1 and 5 of (b) and each other horizontal line (horizontal dissector) by a node of the network. Each rectangle of (a) is represented by a branch connecting the nodes representing the horizontal lines that include the upper and lower sides of the rectangle. The vertical lines (vertical dissectors) in (a) are represented by meshes in (b). Current is considered as flowing downward. If the magnitude of the incoming current is taken as H, the width of the rectangle, then the current in each branch i is equal to the width h_i of rectangle i, the potential drop in i is equal to the depth v_i of rectangle i, and the potential drop between the poles is equal to V. The applicability of the Kirchhoff equations is readily apparent.

No two wires in 2(b) are in series; this would correspond to two rectangles in 2(a) having the entire bottom side of one in common with the entire top side of the other. Likewise, no two wires are in parallel; in this case two rectangles of 2(a) would have a vertical side in common. There is no subnet of more than one wire and less than all wires in series or in parallel with the rest of the net. This two-terminal network is called a p-net (polar net). If the poles are connected by a wire

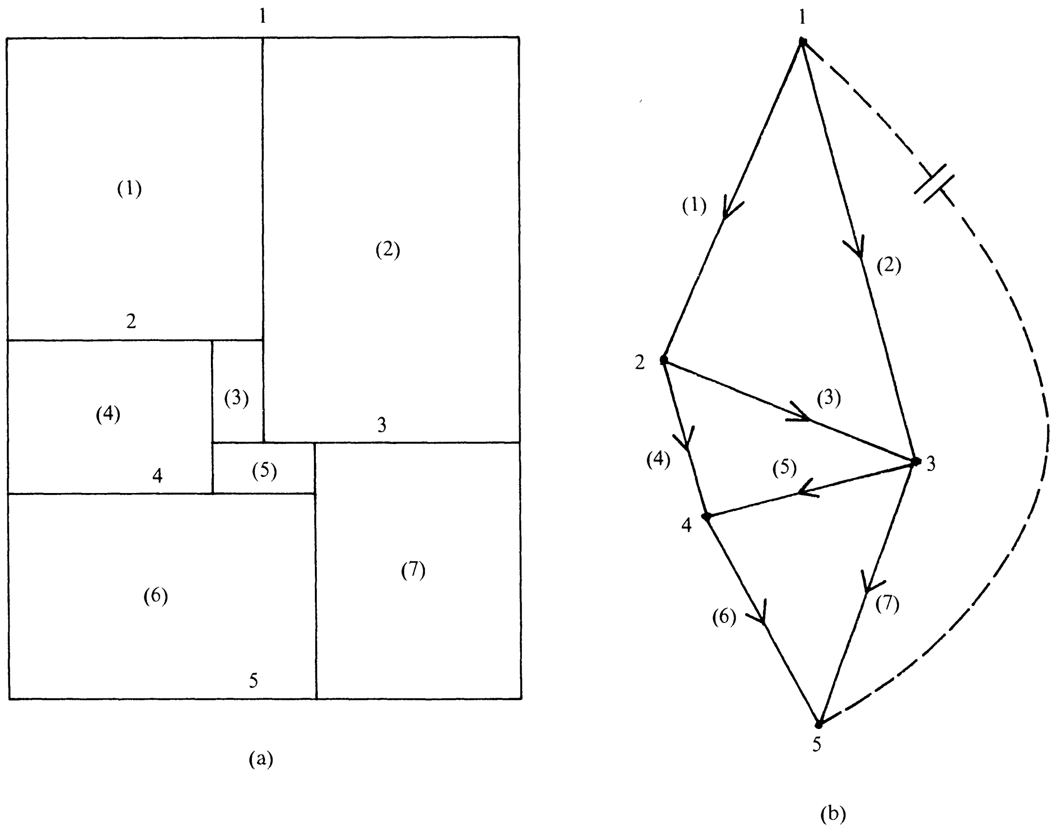


FIG. 2. Rectangled rectangle and net

(the dotted line in 2(b), in which the battery is considered placed) the net with the added wire becomes a 3-connected planar graph. As has been stated, these were called c-nets (completed nets) in the basic paper and in later papers.

Starting with a given c-net with n edges, two adjacent vertices are established as poles and their connecting wire removed (or, equivalently, a battery is considered placed in this wire), and current is considered entering at one pole and leaving at the other. If a rectangle divided into squares is to be produced, the conductance of each wire is taken as unity ($h_i = v_i$) and the Kirchhoff equations for the network are just sufficient to enable the relative numerical values of the currents in the wires to be calculated. These determine a squared rectangle that is fixed to within a scale factor. A c-net with n edges will result in n distinct simple perfect squared rectangles each with $n - 1$ elements, in general. Basic results in [5] are: (1) every simple perfect squared rectangle can be produced from a p-net derived from a c-net, and (2) all the simple perfect squared rectangles of order n are produced from the complete collection of c-nets of order $n + 1$.

If a rectangle divided into squares and one rectangle is to be produced, the conductance of the wires of the p-net is taken to be 1, except for one of them of conductance, say, c. The Kirchhoff equations result in expressions for the relative values of the currents and voltages, linear in terms of the unknown c; but they can be made homogeneous by taking $c = a/b$ and clearing of fractions. The resultant rectangle divided into squares and one rectangle has one degree of freedom. It is variable and has different shapes for different values of c, or a/b. If the width and height of the rectangle are set equal to each other, a numerical value of c is determined and the

result is a square divided into $n - 2$ squares and one rectangle with ratio of horizontal to vertical side equal to c.

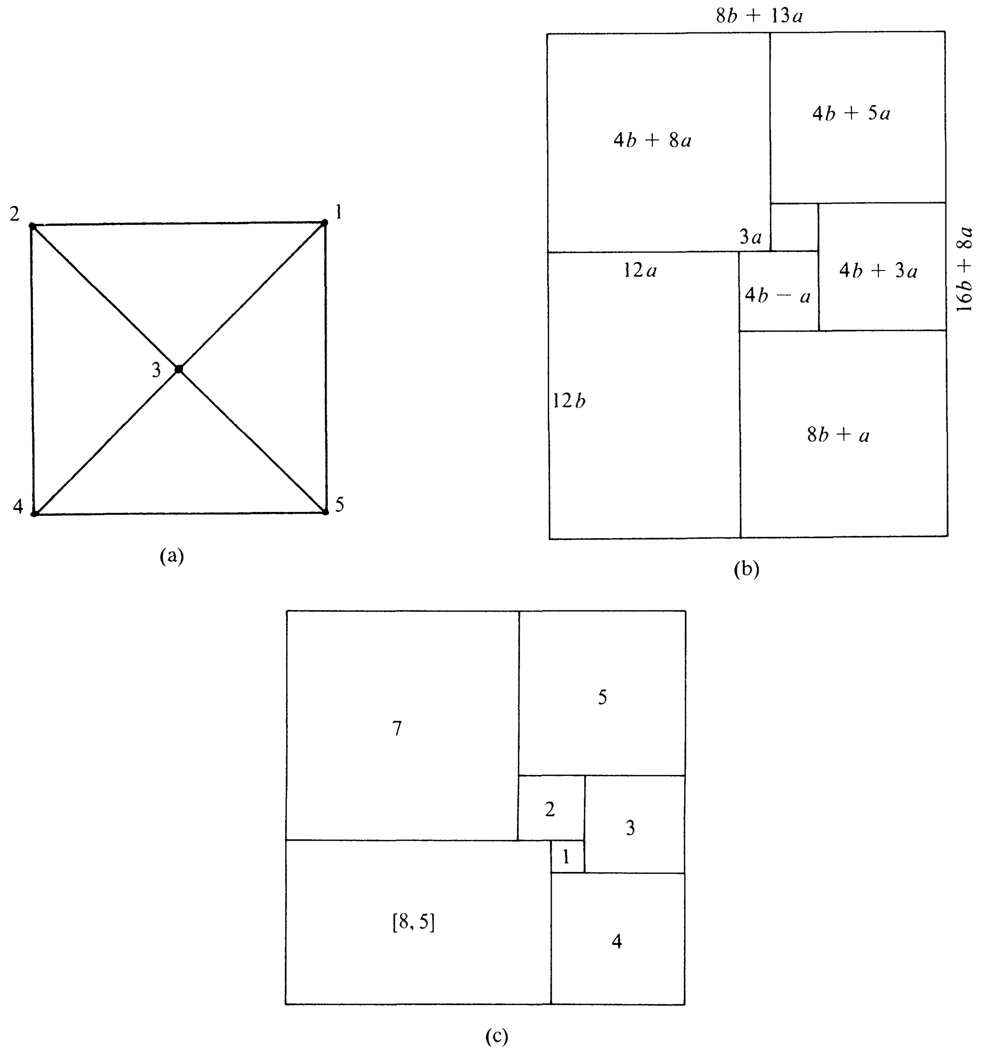


FIG. 3. Derivation of square with one rectangle

C_6, the lowest order c-net, does not result in any D, there is no C_7, and the only C_8 is that shown in Fig. 2(b), with the added wire. This is used as an example to show the derivation and is repeated in different form in Fig. 3(a). Points 1 and 5 are selected as poles and the wire connecting them contains the battery. An arbitrary conductance c is assigned to wire 4-5 and that of the remaining wires taken as unity. The results of solving the Kirchhoff equations, taking $c = a/b$ and clearing of fractions (only relative values are significant at this point), are shown in Fig. 3(b) in the form of the resultant variable rectangle divided into squares and one rectangle.

The shape of the rectangle 3(b) is different for different relative values of a and b (for some values the figure becomes imperfect). For a square, the width $H = 8b + 13a$ must be equal to the height $V = 16b + 8a$. This results in $8b = 5a$ and the relative values are $a = 8$ and $b = 5$

(c = 8/5). The result of using these values, after dividing by a common factor, is shown in Fig. 3(c). This is a square 12 units on each side, divided into 6 unequal squares and a 5 $\times$ 8 rectangle. If a perfect squared rectangle with a ratio of sides 5 to 8 is found, a compound squared square is formed which must then be tested for equal elements. There is in fact a perfect compound squared rectangle of order 18 and sides 115 (= 23 $\cdot$ 5) and 184 (= 23 $\cdot$ 8); these are the sides of the rectangle when the sides of the elements are taken as integers without any common factors. The values in Fig. 3(c) are multiplied by 23 to make the sides of the rectangular space 115 and 184 and the rectangle fitted in. The result, however, is imperfect, as there are two elements with equal sides 115. The rectangle used here is a trivially compound one (see Section 4) and need not even have been tried; a trivially compound rectangle reduced in size to fit into the 5 $\times$ 8 space of 3(c) must have an element of side 5, of which there is already one present.

Referring back to Fig. 3(a), each of the other wires must be assigned the arbitrary conductance and the process repeated; then the battery is to be placed in another branch of the c-net (the p-net can be redrawn so that it lies between the poles), etc. Only two more variable rectangles divided into squares and one rectangle result from this particular c-net. These are shown in Fig. 4. In the first one, setting $H = V$ results in the relative values $a = 2$, $b = 5$, but with these values two of the square elements become equal. In the second one, c is negative, $a = -1$ and $b = 2$, and a square is not possible; there are equal elements as well. These rectangular forms, as well as the one shown in Fig. 3(a), are needed for another purpose and are referred to in Section 4.

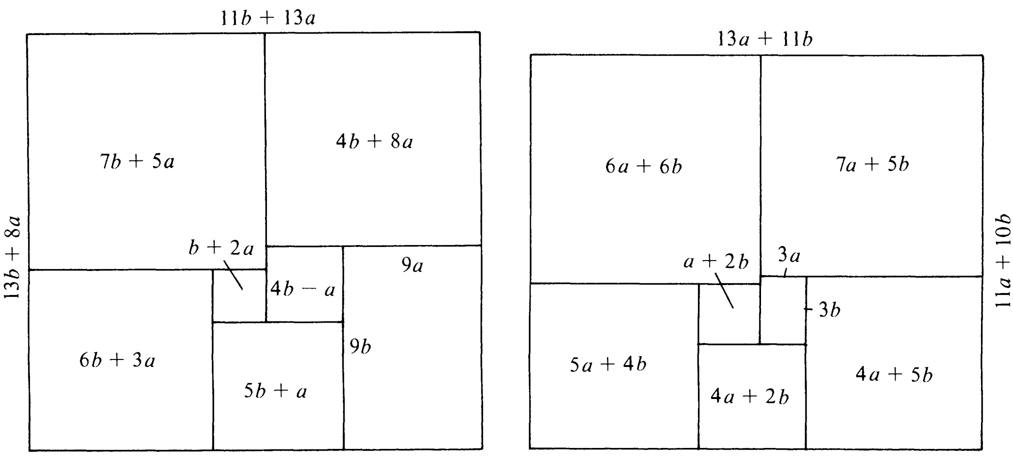


FIG. 4. Squared rectangles with one rectangle

The necessary set of D's with 6 to 15 squares is obtained from the complete set of c-nets (3-connected planar graphs) with 8 to 17 edges, excluding one of each dual pair (duals give the same results rotated 90°). Those from order (number of edges) 8 to 19 (excluding one of each dual pair) have been tabulated and are available [**3**]. These were recently derived again by Duijvestijn for the purpose of obtaining additional information (not relevant here) concerning them. The numbers of c-nets of order 8 to 17, excluding one of each dual pair, range from 1 of order 8 to 671 of order 17, with a total of 1062.

The program for obtaining the D's utilizes theorems from the basic paper [**5**] and electrical network theory to initially obtain a value of $H + V$ and for H in terms of the unknown conductance c placed in one wire, from the topology of the c-net and the p-net. The specific numerical value of this conductance is obtained by setting the value of $H + V$ equal to twice the value for H. As a fraction, the numerator and denominator of c give the relative values of the width (current) and height (voltage drop), respectively, of the rectangular element. Many of the values of c are negative, showing the impossibility of a proper rectangular element. The ratio

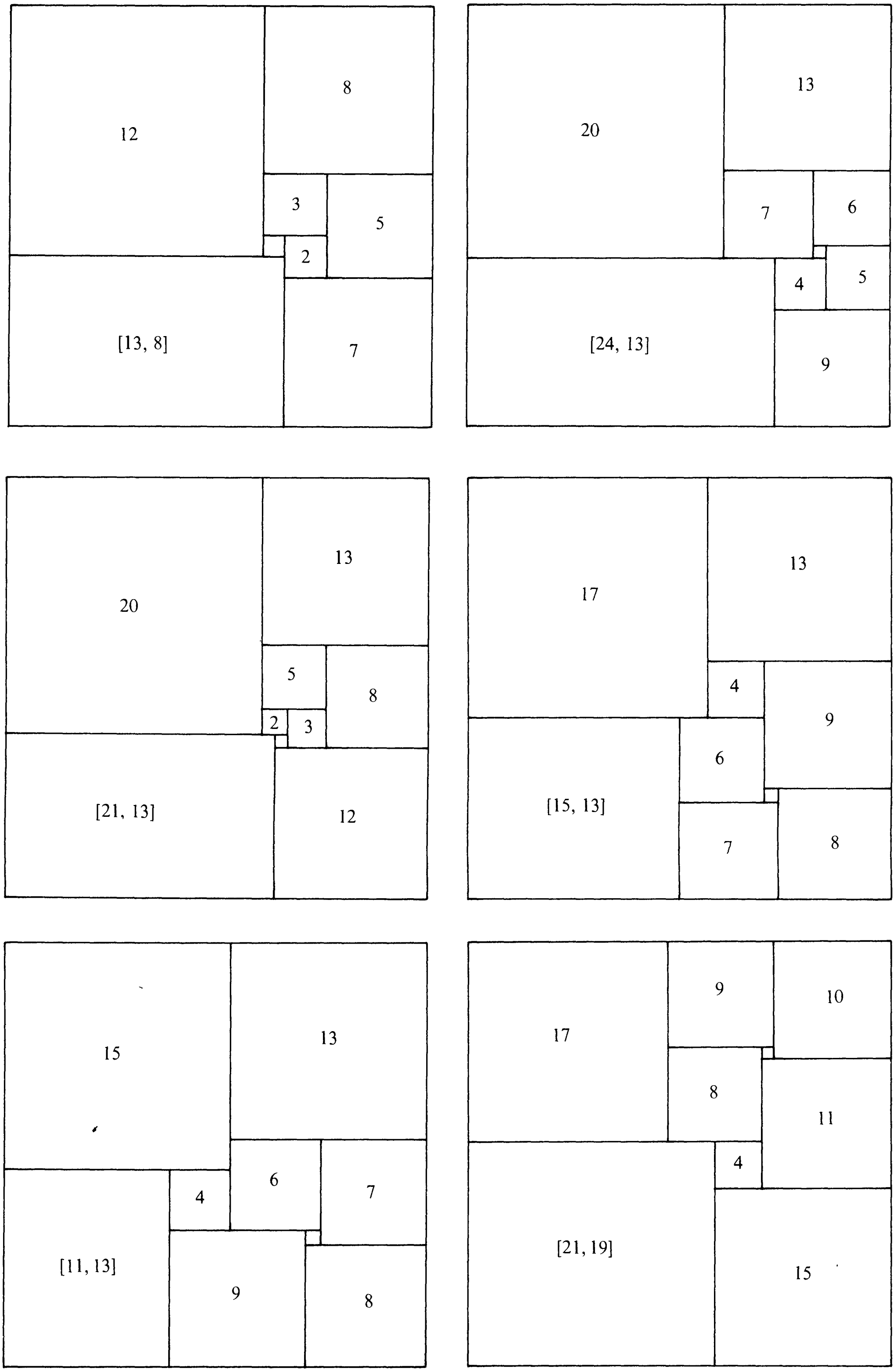


FIG. 5. Some squares with one rectangle

of the short side to the long side of the rectangular element is c if $c < 1$ and $1/c$ if $c > 1$. The square elements for those not eliminated initially were calculated and then tested for imperfections (equal elements).

The single D obtained from C_8 is shown in Fig. 3(c). There are only one from C_9 and five from C_{10}; these six are shown in Fig. 5 in view of the special treatment mentioned in the next section.

The formation of the D's has been described in an expository manner to show the basic principles and so that the method can even be carried out by hand for the simpler cases, as seen by the example used. The actual calculations used for the computer utilize electrical network theory. Details of the program are given in [**13**], and the addendum that follows gives an outline of the current calculations.

4. Squared Rectangles. The perfect squared rectangles to be considered range from order 9 to order 18. The simple perfect rectangles have been derived, from the c-nets up to order 19, and listed by nondecreasing ratio of short side to long side [**4**]. However, they were derived anew for the present program. Those up to order 15 had been listed in the published table [**2**] and those up to order 13 in [**1**].

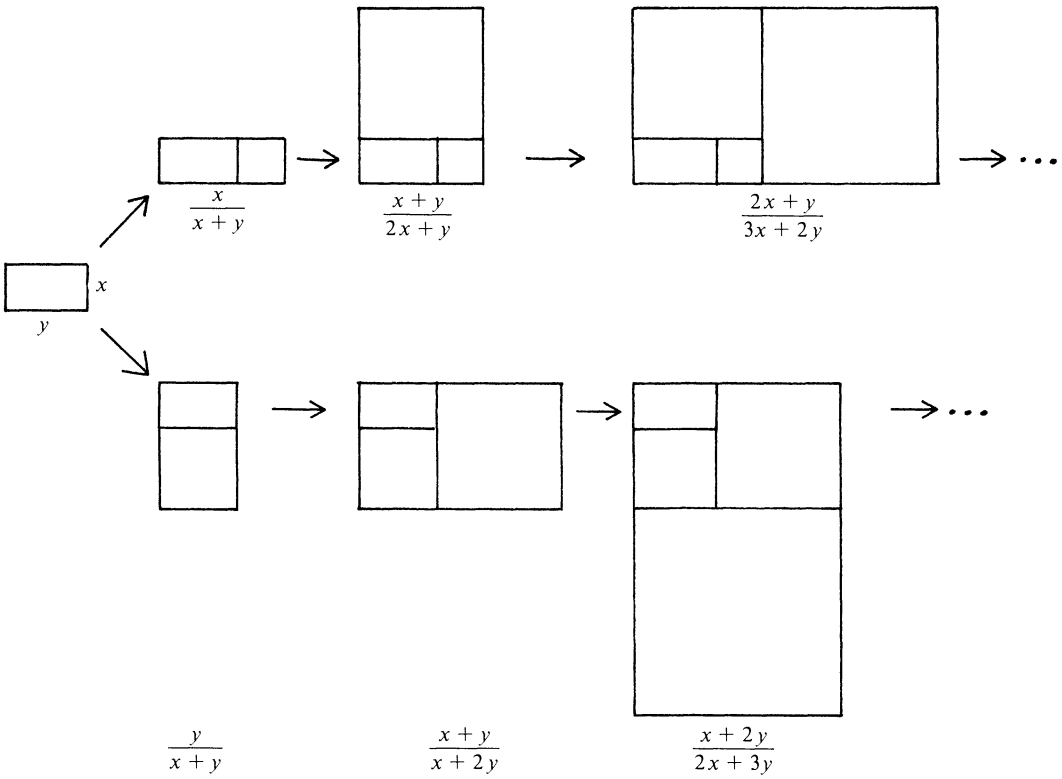


FIG. 6. Trivial compounding

Compound perfect rectangles must also be considered. Compounding is of two kinds, trivial and nontrivial. If a square of side-length equal to a side of a squared rectangle is added along that side, the result is a trivially compound squared rectangle. Two series of these result from each simple squared rectangle as shown in Fig. 6, which also indicates the ratio of sides. (Note that in each series the successive new sides form an additive series and that the ratio of sides approaches τ, the golden ratio, as a limit, in an alternating manner.) While it is necessary to consider the

trivial compounds up to order 18, it is not necessary to actually construct all of them, since the ratio of sides of each is readily expressed in terms of the sides of the basic simple one, and conversely. Only those up to order 15 were constructed. This eliminated the addition of over 120,000 rectangles to the list. For a complete canvass up to order 24 squares, only trivial compounds of order 16, 17, and 18 are needed for D_6, D_7 and D_8. As noted in Section 3, there are only seven of these, illustrated in Figs. 3(c) and 4. With respect to five of them, any trivial compound rectangle that fitted would necessarily introduce an element the same size as one already present. For the remaining two, any trivial compound rectangle that fitted the 11×13 space of one or the 19×21 space of the other would require, working backwards, base simple or nontrivial compound rectangles of ratio of sides $2/11$ or $2/9$ for one and $2/19$ or $2/17$ for the other. Such rectangles do not exist up to order 18.

In the description of the formation of the D's in Section 3, the first step was the construction of rectangles divided into squares and one rectangle. The three obtainable from C_8, which each contain 7 squares and one rectangle, are shown in Figs. 3(b) and 4. These are *Frames* for the construction of nontrivial compound rectangles. Each perfect squared rectangle from order 9 to order 12 (compound as well as simple ones) is to be fitted into the rectangular space of each of these Frames. This is done by setting the two sides of the rectangle equal to the two sides of the rectangular space, thus fixing the relative values of the variables. This is done in two ways for each rectangle. Nontrivial compounds from order 15 to 18 are produced from these particular Frames; some may be imperfect and are weeded out.

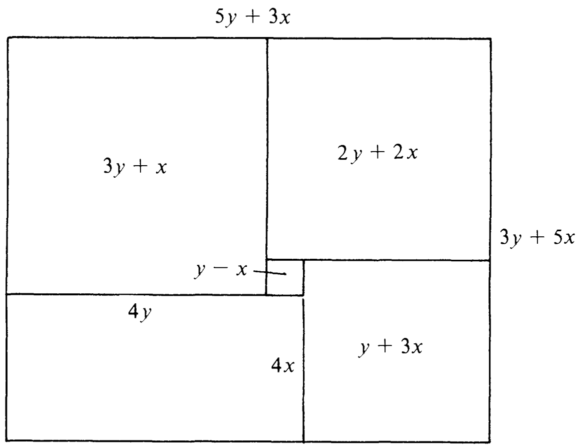


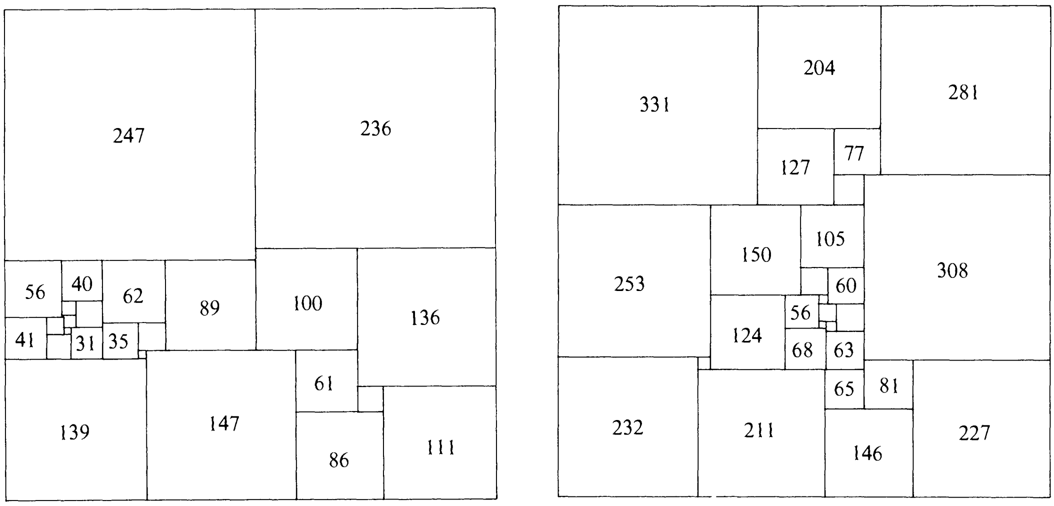
FIG. 7. Example of a Frame

Since the lowest-order perfect rectangle is 9, Frames with up to 9 squares are needed to reach 18 squares. These are derived from c-nets up to order 11, in the manner described in Section 3. The lowest order c-net, C_6, gives a Frame with 4 squares, shown in Fig. 7. For this one, squared rectangles from order 9 to 14 (compound as well as simple) are used, giving nontrivial compounds from order 13 to 18. The nontrivial compounds below order 15 are also further compounded (trivially or nontrivially) in any manner or combination applicable, producing some additional nontrivial compounds up to order 18, as well as additional trivial compounds up to order 15.

Finally, compound rectangles containing two subrectangles are also to be considered. Up to order 18, these can only result from the juxtaposition of two rectangles of order 9 (of which there are only two), in several ways. Each, which may be the same rectangle rotated 90°, is multiplied by the appropriate factor to make a side of one equal to a side of the other. There were only a half-dozen of these, and they were considered individually.

5. Results for Type 1 Squares. The remainder of the program relating to type 1 squares involved matching the ratio of the rectangular spaces in the D's with the ratio of sides of the rectangles, with the additional individual treatment of the low-order D's already noted. Each matching pair gave a compound squared square that then needed to be checked for the presence of equal elements.

The basic results were that there are no compound perfect squares below order 24 and only the known single one of order 24, shown in Fig. 1(a).



(a) 26-483 (b) 28-816

FIG. 8. Some new compound squares

The number of type 1 perfect squares produced was 1998, of which 1883 are new. Two of these are shown in Fig. 8. The forms shown are not common; other forms, more complicated and more uncommon, cannot be shown, as the scale would be too small. The subrectangle is a simple squared rectangle located in a corner of the square in 84 percent of the cases; altogether, 90 percent are in a corner and 94 percent of all the subrectangles are simple.

Table 1 presents a count of the output, arranged according to the order of the squares and the order of the D's. The order of the subrectangle for any particular group can of course be obtained by subtraction. Numbers in italics are in those combinations of D's and rectangles that were not completely canvassed.

The first line of totals in Table 1 gives the output according to the order of the squares. Some were already known; these are noted in the next line of the table. The differences give the number of new squares, according to their order. The next line of the table gives the number of known squares that were outside the ambit of the program, and the last line gives the overall totals.

The number of previously known type 1 squares from order 24 to 33 was 235. Of these, 115 were of such a nature that they should have been produced by the program; all of them were, and this served as a check. The remainder, 120 in number ranging from one of order 27 to 50 of order 33, were outside the scope of the program as carried out; some (102) would come from a D higher than D_{15}—for example, the square of Fig. 1(e) would come from a D_{17}; some (6) have a subrectangle of order 19 or higher; and the remainder have a trivially compound subrectangle of order 16 or higher not included in part of the program.

6. Type 2 Compound Squares. The treatment of type 2 compound squares is divided into two parts. The first part concerns squares made up by the simple juxtaposition of two properly

TABLE 1
Results for Type 1 Squares

D	24	25	26	27	28	29	30	31	32	33	Total
6	0										0
7	0	0									0
8	0	0	2								2
9	0	0	1	2							3
10	0	0	2	1	4						7
11	1	0	2	1	2	15					21
12	0	1	0	5	3	8	42				59
13	0	1	2	3	8	13	32	86			145
14	0	0	1	8	9	29	46	131	214		438
15	0	0	2	5	21	74	68	91	294	768	1323
Total (1)[a]	1	2	12	25	47	139	188	308	508	768	1998
Old (in)[b]	1	2	10	18	24	38	16	1	0	5	115
New[c]	0	0	2	7	23	101	172	307	519	763	1883
Old (out)[d]	0	0	0	1	9	11	3	38	8	50	120
Total (2)[e]	1	2	12	26	56	150	191	346	516	818	2118

(a) Results of program

(b) Squares in Total (1) already known

(c) Difference

(d) Known squares outside scope of program

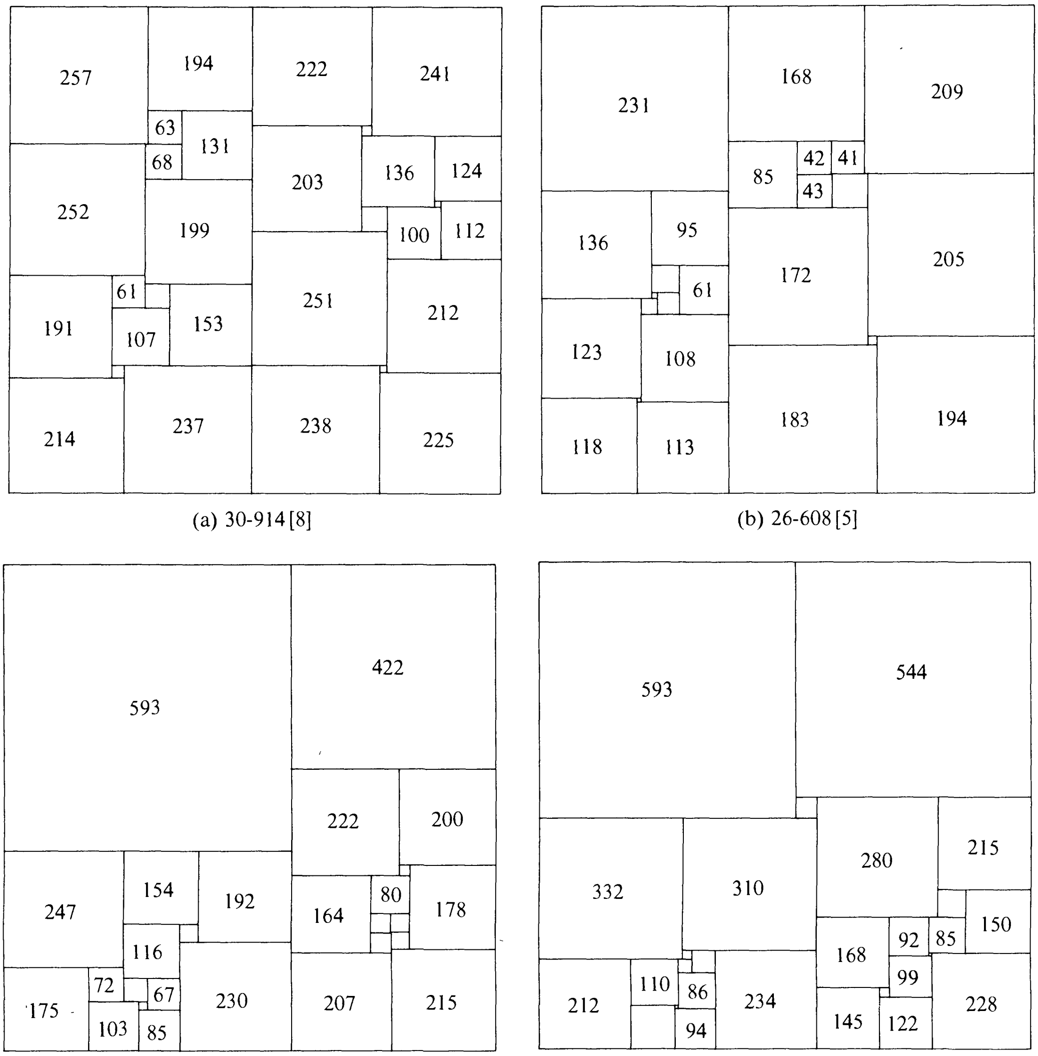
(e) Total squares now known

proportioned rectangles. Fig. 9 shows the three kinds. In (a), neither rectangle is a trivial compound. In (b), one of the two is a trivial compound. In (c), both rectangles are trivially compound. (The fourth square (d) relates to the second part.) Since the lowest-order rectangle is order 9, only rectangles up to order 15 (simple, trivial compound, and nontrivial compound) are needed for a complete canvass up to order 24 squares, but the full list of rectangles used for type 1 squares was also used here. The method consists in finding two rectangles for which the ratios of short side to long side sum to one.

Other type 2 squares are formed by first constructing squares divided into unequal squares and two rectangular spaces, called DD's (doubly deficient squared squares); one is shown in Fig. 9(d). These are formed from c-nets in a manner similar to the D's (Section 3), except that two wires of the network are given arbitrary conductances in all possible ways, instead of just one. The DD with the fewest number of unequal squares contains 3 squares; hence only rectangles up to order 12 are required for a complete canvass up to order 24. Two rectangles of order 9 need 6 more squares to reach 24; hence only the DD's up to 6 squares are required. These were formed from the C_6, C_8, and C_9 c-nets, of which there are only 3. Since DD's with 7 or more squares were not used, the canvass was not complete for order 25 and higher.

The condition that $H = V$ imposes a relationship between the two arbitrary conductances. The elements of the DD then can be worked out as linear expressions in terms of two homogeneous variables.

Fig. 10 shows at (a) the single distinct DD that results from C_6 and at (b) one of those obtained from C_8. Each rectangle from the stock is fitted into one of the two rectangular spaces (which must be done in two ways); this fixes the relative values of the two variables and hence also fixes the relative values of the sides of the other rectangular space. The result is a squared square (compound) with one rectangular space. Then a squared rectangle is sought, in the same manner as for type 1 squares, having the same ratio of sides as the second rectangular space. The program



(a) 30-914[8]

(b) 26-608[5]

(c) 28-1015[17]

(d) 28-1137[8]

FIG. 9. Some known type-2 squares

as carried out used only the rectangles (simple and compound) of orders 9 to 12 for the first rectangle but then searched the entire list for the second rectangle.

The usable networks were also derived by considering all the connected graphs with from 2 to 8 edges and eliminating the impossible ones, that is, those that could not possibly lead to a two-terminal net having the required properties. The results were the same as those obtained from the c-nets, with the addition of the two-wire net, each wire having arbitrary conductance, corresponding to the juxtaposition of two rectangles.

The results of this part showed that there are no type 2 squares of order 24 or lower. This field had already been pretty well worked over, and no new squares below order 30 were found. Table 2 gives the number of type 2 squares, including old ones outside the scope of the program being carried out. The four groups listed follow in order the four kinds described, and the totals have the same significance as in Table 1.

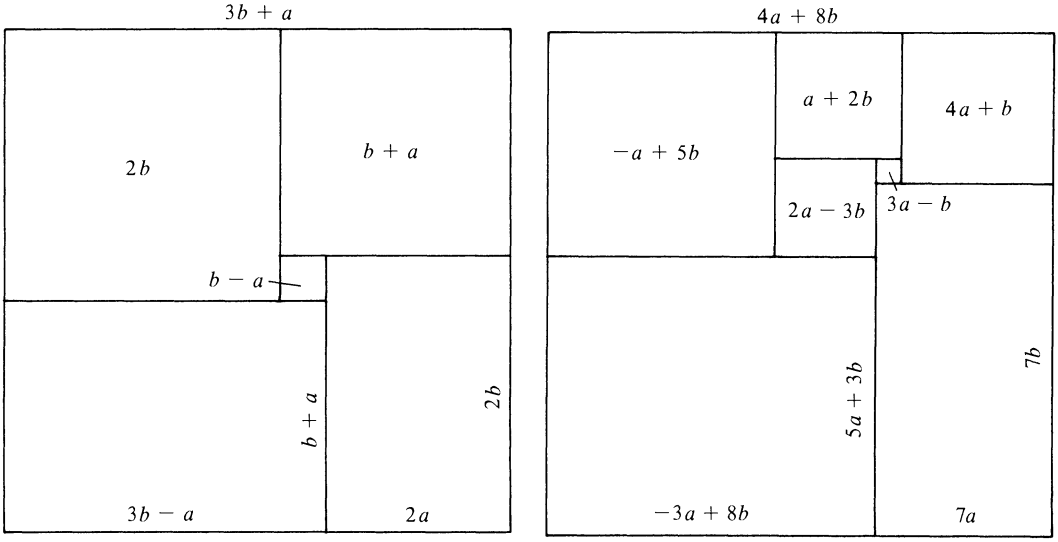


FIG. 10. Examples of squares with two rectangles

TABLE 2
Results for Type 2 Squares

	24	25	26	27	28	29	30	31	32	33	Total
(a)	0	0	0	0	0	0	1	0	0	0	1
(b)	0	0	1	0	2	1	6	1	8	4	23
(c)	0	0	0	0	1	0	2	0	0	0	3
(d)	0	0	0	0	1	0	3	13	7	23	47
Total (1)	0	0	1	0	4	1	12	14	15	27	74
Old (in)	0	0	1	0	4	1	9	0	0	0	15
New	0	0	0	0	0	0	3	14	15	27	59
Old (out)	0	0	0	0	0	0	5	1	10	3	19
Total (2)	0	0	1	0	4	1	17	15	25	30	93

TABLE 3
Number of Compound Perfect Squares Found

Order	Number	New
24	1	0
25	2	0
26	13	2
27	26	7
28	60	23
29	151	101
30	208	175
31	361	321
32	541	523
33	848	790
Total	2211	1942

7. Summary and General Comments. Table 3 gives a summary of the total number of compound perfect squares to order 33 now known. There are none below order 24 and only one of order 24. The number for order 25 is probably only 2, but this has not been shown. The canvass for order 25 and higher was not complete, with the incompleteness increasing with the order.

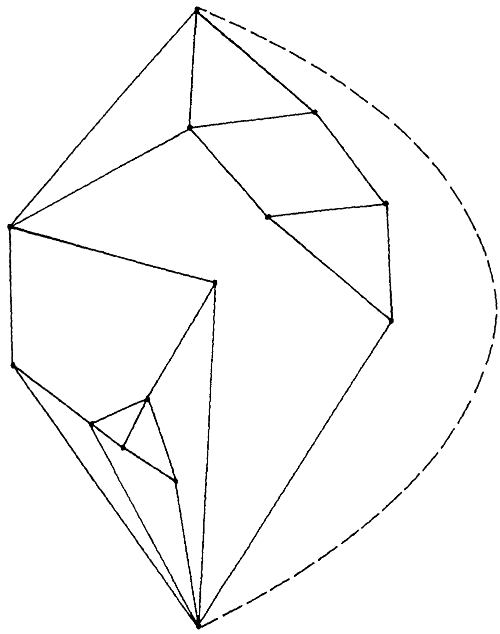


FIG. 11. Net of a compound square

If a compound squared square is represented by a two-pole network, the result will have at least 24 wires, and the addition of the wire to complete the net will result in a minimum of 25. Fig. 11 is the network of the 24-order compound square shown in Fig. 1(a). With the added wire, shown by a dotted line, it has 25 wires. A direct approach to the problem of determining whether there are any compound squares below order 24 would be to construct all the nets of the appropriate type, one of which is shown in Fig. 11, with up to 24 wires. This was the approach followed by Kazarinoff and Weitzenkamp [**10**], [**11**]. They defined the structure of this type of net in graph theory terminology and constructed those with 22 or fewer wires. After certain *a priori* eliminations the result was a collection of over 17,000 nets. These then were treated (by computer) one by one; this would involve placing a battery in one edge (each edge, in turn, presumably with some *a priori* eliminations) and solving the Kirchhoff equations, the conductance of each wire being unity. The result was that no compound perfect squares with 21 or fewer elements were found. They also derived [**11**] simple perfect rectangles to order 17 and higher and gave some tables concerning them. Their Table II, a list of rectangles up to order 18 with ratio of sides p/q, where $p + q < 30$, has an omission. [The missing one, order 18 and $p/q = 9/10$ has the following code of its elements: (163, 127) (55, 72) (98, 46, 19) (21, 36, 17) (27, 62) (6, 15) (52) (43, 8) (35); these are read from left to right within each set of parentheses, the groups are in order of decreasing height.]

The completed net of Fig. 11 has connectivity 2. It has a part, with more than one wire and less than all, connected to the rest of the graph at only two points. The subgraph representing the subrectangle is connected to the rest of the graph at two points only. The method used here replaces this subgraph by a single wire not having unit conductance. In the example, the 13 wires representing the subrectangle are replaced by a single wire (of conductance 111/94) and the

25-order graph of connectivity 2 is reduced to a 3-connected graph with only 13 wires. The number of graphs needing treatment were 1062 to produce the D's, 7 to produce the Frames for compound rectangles, and 3 to produce the DD's. The method is based on the obvious extension of results in the basic paper [**5**] relating to simple squared rectangles, to simple rectangled rectangles.

Addendum. Following is a brief outline relating to the current calculations.

The number of vertices of a c-net is denoted by V, the number of faces by F, and the number of edges by E. The vertices are numbered, in any order, from 1 to V. A face is coded by listing the sequence of its vertices in a positive sense; that is, the boundary is traversed keeping the inside of the face on the left. Any vertex can be the starting point, and this point is repeated at the end. A code of the net is a sequence of the codes of its faces, separated by zeros and with two zeros at the end. Thus a code of the c-net of Fig. 3(a) is:

432405345051350312301542100.

This code embodies the planarity of the c-net. A list of the edges and the code of the dual net can be readily obtained from it [**6**]. An edge is determined by its endpoints and is represented by two one-dimensional vectors; br1[i] and br2[i], $i = 1(1)E$, represents edge i and br1 and br2 are arrays of the vertices. From the knowledge of the code only, it is possible to obtain the current distribution in the branches of the c-net considered as an electrical network when electromotive forces are placed in branches.

The special vertex-vertex matrix INC, defined as follows,

$$\begin{aligned} \mathrm{INC}\,[i,j] = \mathrm{INC}\,[j,i] &= -1 \quad \text{if } i \text{ and } j \text{ are adjacent,} \\ &= 0 \quad \text{if } i \text{ and } j \text{ are not adjacent,} \\ \mathrm{INC}\,[i,i] \qquad\qquad\quad &= \text{valence of vertex } i, \end{aligned}$$

can readily be obtained from the representation of the edges. This matrix is symmetric but singular. From the basic paper [**5**] it is known that all its first cofactors are equal; their common value is called the complexity of the c-net and is denoted by C. Its value is equal to the semi-perimeter of any rectangle derived from the c-net, and the complexity of the p-net, the c-net with one branch disregarded, is equal to the breadth of the corresponding rectangle.

A current source IV[i] is assumed imposed on vertex i. The branch currents IB and the current sources IV are connected by the relation IV = A × IB, where IB is a vector of the branch currents having E elements and A the vertex-branch incidence matrix having V rows and E columns. Furthermore we define the vector EB with elements EB[i] denoting the voltage difference over branch br1[i], br2[i] and the vector EV of voltages of the vertices, where EV[i] denotes the voltage of vertex i. The voltage EV[V] is taken as zero. According to Kron and Cauer [**14**], [**15**], the following relation holds: EB = A$'$ × EV, where A$'$ is the transpose of A.

A conductance matrix G is defined, where G[k, k] denotes the conductance in branch br1[k], br2[k]. The following relations hold:

$$\mathrm{IB} = \mathrm{G} \times \mathrm{EB} \quad \text{and} \quad \mathrm{IV} = \mathrm{A} \times \mathrm{IB} = \mathrm{A} \times \mathrm{G} \times \mathrm{EB} = \mathrm{A} \times \mathrm{G} \times \mathrm{A}' \times \mathrm{EV}.$$

Let $\eta = \mathrm{A} \times \mathrm{G} \times \mathrm{A}'$. If G is the unit matrix then $\eta = \mathrm{A} \times \mathrm{A}'$ is equal to the matrix INC. Since η is singular and all its first cofactors are equal, the number of equations must be reduced by one. Instead of using the vectors IV and EV, the last element of each is discarded to form the reduced vectors IVR and EVR. The matrix α is defined by the relation IVR = α × IB. Obviously α can be obtained by omitting the last row from A. Furthermore EB = α' × EVR. The reduced matrix ηR can be obtained from A × G × A$'$ by removing the last row and last column; it is nonsingular. It can be shown that

$$\mathrm{EB} = \alpha' \times (\alpha \times \alpha')^{-1} \times \mathrm{IVR}.$$

This means that given the source currents the voltages over the branches can be calculated.

Consider one branch with vertices br1[m] and br2[m]. Set IVR[br1[m]] equal to C and IVR[br2[m]] equal to $-$C, where C is the complexity of the c-net, and all other elements of IVR equal to zero. The situation at the branch is shown in Fig. 12 at the left, where b is the current in the branch. According to Thevenin's theorem this is equivalent to an EMF in that branch causing a current $C - b$ in the same branch. This means that

$$\mathrm{S} = \mathrm{C} \times \alpha' \times (\alpha \times \alpha')^{-1} \times \alpha - \mathrm{C} \times \epsilon,$$

where ϵ is the unit matrix, describes the current flow in the network. S is a symmetric matrix with E rows and E columns. An element S[m, w] denotes the current through branch br1[m], br2[m] caused by an EMF of value $-$C in branch br1[w], br2[w]. In testing for a squared square solution only the elements in the main diagonal of S are required and only if the current through the EMF branch is half the complexity need the solution be considered. In case a conductance c, other than one, is placed in a branch, all currents are linear combinations of c. It is therefore sufficient to calculate the currents taking $c = 1$ and $c = 0$. Two calculations are performed, one with the conductance matrix G with all conductances equal to 1 and another in which the conductance of a particular branch is replaced by zero.

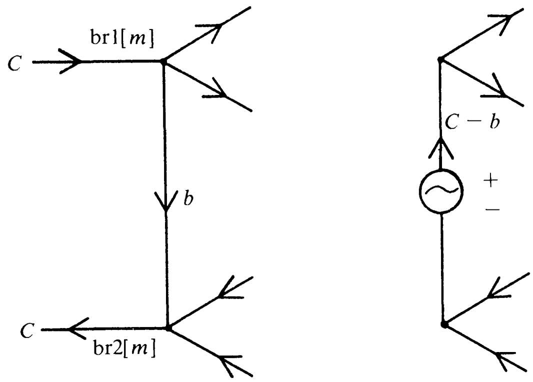


FIG. 12.

Full details of the method are in [**6**], [**13**], and [**16**], in which references the various programs are also given.

References

1. C. J. Bouwkamp, On the dissection of rectangles into squares I, Koninkl. Nederl. Akad. Wetensch. Proc. Ser. A, 49 (1946) 1172–1178; II, III, 50 (1947) 58–71, 72–78.

2. C. J. Bouwkamp, A. J. W. Duijvestijn, P. Medema, Tables relating to simple squared rectangles of orders 9 through 15, Technische Hogeschool, Eindhoven, Netherlands, 1960.

3. ______, Table of c-nets of orders 8-19, inclusive, Philips Research Laboratories, Eindhoven, Netherlands, 1960; unpublished, available in UMT file of Mathematics of Computation.

4. C. J. Bouwkamp, A. J. W. Duijvestijn and J. Haubrich, Catalogue of simple perfect rectangles of orders 9 through 18, Philips Research Laboratories, Eindhoven, Netherlands, 1964; unpublished.

5. R. L. Brooks, C. A. B. Smith, A. H. Stone, and W. T. Tutte, The dissection of rectangles into squares, Duke Math. J., 7 (1940) 312–340.

6. A. J. W. Duijvestijn, Electronic computation of squared rectangles, dissertation, Technische Hogeschool, Eindhoven, Netherlands; also in Philips Res. Rep., 17 (1962) 523–612.

7. ______, Simple perfect squared square of lowest order, J. Comb. Theory, B25 (1978) 240–243.

8. P. J. Federico, Note on some low-order perfect squared squares, Can. J. Math., 15 (1963) 350–362.

9. ______, Squaring rectangles and squares: A historical review with annotated bibliography, in Graph Theory and Related Topics, G. A. Bondy and U. S. R. Murty, eds., Academic Press, 1979, pages 173–196.

10. N. D. Kazarinoff and R. Weitzenkamp, On existence of compound perfect squares of small order, J. Comb. Theory, B14 (1973) 163–179.

11. ______, Squaring rectangles and squares, this MONTHLY, 80 (1973) 877–888.

12. T. H. Willcocks, Problem 7795 and solution, Fairy Chess Review, 7 (1948) 97, 106.

13. P. Leeuw, Electronic computation of compound squared squares, Bachelor's thesis, Twente Technical University, Enschede, Netherlands, Jan. 1979.

14. W. Cauer, Theorie der linearen Wechselstromschaltungen, 2nd ed., Akademie Verlag, Berlin, 1954, pp. 56–91.

15. G. Kron, Tensor Analysis of Electrical Networks, Wiley, New York, 1939.

16. A. J. W. Duijvestijn, An algorithm that investigates the planarity of a network, in Computers in Mathematical Research, R. F. Churchouse, C. J. Herr, eds., North Holland, Amsterdam, 1968, pp. 71–72.

17. A. H. Stone, problem E401, this MONTHLY, 47 (1940) 48.

18. E. Lainez, personal communication, 1970.

19. T. H. Willcocks, personal communication, 1964.

MISCELLANEA

66.

RETROACTIVE EDITORIAL POLICY

R. P. BOAS

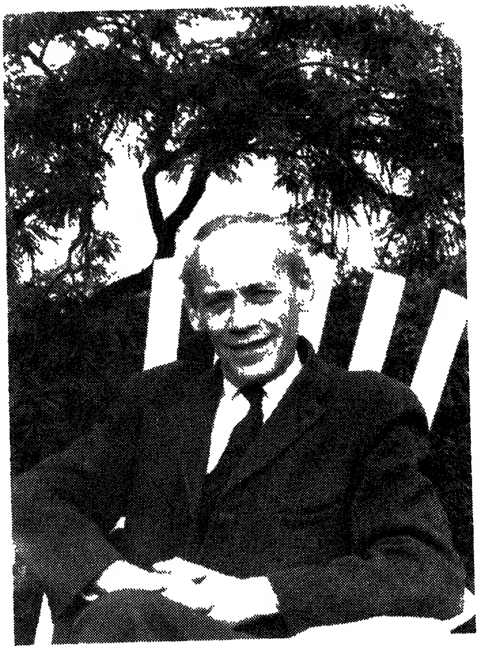
"Let me make it clear to you
This is what we'll never do."

It really doesn't matter if you don't know how to spell:
You'll find that many readers understand you just as well;
And once the spelling's gone to pot, why then I rather guess
It doesn't really matter if the syntax is a mess;
    But not in *my* journal.

You really shouldn't worry that it's all been said before,
Since checking out the sources can be an awful bore.
And don't be very troubled if you seem to plagiarize:
If you copy something really good, you just might win a prize;
    But not in *my* journal.

We often note that authors, even those whose work is strong,
They sometimes go too far and say a thing or two that's wrong.
You needn't worry very much about a stray mistake:
If you can fool the referee, what difference does it make?
    But not in *my* journal.

Oh, inspiration's wonderful but second thoughts are best,
And don't you think we must have made those silly rules in jest?
You seekers for perfection who've made an utter goof,
You'll have a chance to fix it up: rewrite in galley proof;
    But not in *my* journal.



Since, as is well known, all mathematics is divided into four parts, a representative collection of mathematicians' photographs should contain an algebraist, a geometer, an analyst, and an applied mathematician. Here, but not necessarily in that order, is such a collection—in fact the order is alphabetical. The secret is revealed on p. 58.

$$¿ \quad \operatorname{per}(A \circ \bar{A}) \leq (\operatorname{per} A)^2. \quad ? \tag{5}$$

Now assume the truth of (5) for all positive semidefinite hermitian matrices A. Note that

$$\operatorname{per}(A \circ \bar{A}) = \sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}|^2.$$

Then,

$$0 \leq \operatorname{per}(A \circ B) = \sum_{\sigma \in S_m} \prod_{i=1}^{m} a_{i\sigma(i)} b_{i\sigma(i)} \leq \sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}| |b_{i\sigma(i)}|$$

$$\leq \left(\sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}|^2 \right)^{1/2} \left(\sum_{\sigma \in S_m} \prod_{i=1}^{m} |b_{i\sigma(i)}|^2 \right)^{1/2} = \left(\operatorname{per}(A \circ \bar{A})\right)^{1/2} \left(\operatorname{per}(B \circ \bar{B})\right)^{1/2},$$

$$\operatorname{per}(A \circ B) \leq \operatorname{per} A \operatorname{per} B.$$

To get the second line, we applied the Cauchy-Schwarz inequality. Therefore, inequalities (4) and (5) are equivalent.

4. Example. If

$$A = \begin{bmatrix} 2 & -i & 1 \\ i & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 3 & 2i \\ 3 & 6 & i \\ -2i & -i & 5 \end{bmatrix},$$

then

$$A \circ B = \begin{bmatrix} 8 & -3i & 2i \\ 3i & 12 & 0 \\ -2i & 0 & 15 \end{bmatrix}.$$

It is clear that all 1×1, 2×2 and 3×3 principal subdeterminants of the hermitian matrices A, B and $A \circ B$ are nonnegative. Therefore, by a known theorem, these matrices are positive semidefinite hermitian. A short calculation now shows that $\operatorname{per} A = 17$, $\operatorname{per} B = 205$, $\operatorname{per}(A \circ B) = 1623$ and $\operatorname{per} A \operatorname{per} B = 3485$.

References

1. M. Marcus, The Hadamard theorem for permanents, Proc. Amer. Math. Soc., 15 (1964) 967–973.
2. M. Marcus and H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber and Schmidt, Boston, 1964.
3. H. Minc, Permanents, Addison-Wesley, Reading, Mass., 1978.
4. A. Oppenheim, Inequalities connected with definite hermitian forms, J. London Math. Soc., 5 (1930) 114–119.
5. G. P. H. Styan, Hadamard products and multivariate statistical analysis, Linear Algebra and Appl., 6 (1973) 217–240.

ANSWERS TO "PHOTOS" ON PAGE 33

Top left: E. Artin; top right: S. Banach; bottom left: A. Grothendieck; bottom right: C. Shannon. Artin, Banach, and Shannon would probably be recognized by their friends, but Grothendieck probably not. Grothendieck's picture was taken in 1954 ± 1.

COMPUTERS AND THE MULTIPLICITY OF POLYNOMIAL ROOTS

JOHN J. WAVRIK
Department of Mathematics, University of California, San Diego, La Jolla, CA 92093

Introduction. This article will not describe the assorted twists and turns of fate that lead a worker in a pure mathematics area like algebraic geometry to become involved with computers. It is quite likely, however, that as personal computers become more common an algebraist who acquires one will at some stage make a stab at using it for research work. This article, written by an algebraist who has become so involved, will try to prepare others for the shock, amazement, problems, pitfalls, and pleasures which lie in the wonderful world of machine computation.

Algebraic geometry is an example of a field that has developed in a way which underemphasizes computation. Recent generations of students have been trained to accept an outlook on the subject in which even the geometry aspects have been all but hidden from view. Andre Weil comments upon this in the introduction to *Foundations of Algebraic Geometry*:

> Nor should one forget, when discussing such subjects as algebraic geometry, and in particular the work of the Italian School, that the so-called "intuition" of earlier mathematicians, reckless as their use of it may sometimes appear to us, often rested on a most painstaking study of numerous special examples, from which they gained an insight not often found among modern exponents of the axiomatic creed.

The ability to perform explicit computation and examine the details of examples provides guidance and intuition in even the most theoretical of fields.

Weil uses the word "painstaking." Those who have tried to work out details of examples might add adjectives like "time-consuming," "arduous," and "laborious." It is natural to hope that computers can be used to reduce the difficulty of computation. As it turns out, the process is not as simple as it might first appear.

Algebraic geometry studies the set of zeros of polynomial equations. If $f(x, y)$ is a polynomial in two variables, the set of points (x, y) in the real plane which satisfy $f(x, y) = 0$ will usually look like a curve. As the subject developed it appeared both useful and interesting to allow the coordinates of points to lie in fields other than $\mathbb{R}$ and to study "points at infinity" by looking at points in projective spaces. Gradually further levels of abstractions and generality were added. In any event, we have some objects of interest to use and, as in many other branches of mathematics, the objects are often studied by attaching "invariants" to them. (In the past, "invariants" were numbers—now they may be algebraic objects: groups, rings, vector spaces, etc.) We would like to have machine assistance, for example, in computing invariants for specific instances of the objects of study.

Numerical invariants can often be defined as the multiplicity of a root of a polynomial; so we will consider the problem of determining the multiplicity of polynomial roots by machine computation.

The purpose of this article is to describe the stages in the development of a computer program to solve a particular problem in algebra. It is addressed to readers whose background is in mathematics rather than computer science. It is hoped that this article will prove useful even to readers not interested in the specific problem solved by the program.

1. Computers and Languages. A computer is a device which moves electrical signals about at a

The author is an associate professor at the University of California at San Diego. He did his graduate work as a student of Kunihiko Kodaira in the area of deformations of complex structures and completed his Ph.D. in 1966 at Stanford University. He became interested in the potential uses of computers in pure mathematics when he first used one to solve a mathematical problem five years ago. Since that time he has been continuing research in applications of computers, and particularly microcomputers, to pure mathematics. He is also interested in education. Currently he and his wife Mary work in schools to explore uses of microcomputers in education.—*Editors*

rapid rate. Most users of computers never become involved with the machine on this level—system designers and language designers have arranged things so that the internal aspects of a computer (the actual movement of electrical signals, the storage and representation of "information") are "transparent" to most users. What most users see is a device that seems to operate (quite rapidly) with whatever type of information is of concern to them. A hand calculator, for example, seems to be a machine which operates with numbers. The way in which numbers are represented internally and the way the representations are manipulated are not visible to the user. It is easy, therefore, for users to be completely ignorant of what is going on "inside" in even the most general way. In many cases ignorance is, in fact, bliss. The fact that computers can be made to conceal these details from the user contributes to their widespread usefulness. Most users do not so much interact with a machine as with a language. "Higher level" computer languages (FORTRAN, ALGOL, BASIC, Pascal, etc.) are more highly stylized and less expressive than natural languages, but they are a good bit closer to natural language than "lower level" machine languages. A computer language permits a limited range of discourse in an area of interest to the intended users.

Computers were originally developed to facilitate complicated and time-consuming computations arising in science and engineering. The early "higher level" languages were designed with these uses in mind: Most scientific computations involve the real numbers, and so languages were developed to deal with the arithmetic of the real numbers.

Computers are finite machines, and so real numbers are represented in scientific notation: a certain number of "significant digits" together with an "exponent." A natural first step is to try using "floating-point" arithmetic to solve the problem.

We will henceforth use the word "machine" to refer to a combination of computer, operating system, and language. Our discussion, then, concerns the virtual machines with which most users interact.

1.1. At first glance the problem appears quite simple: apply standard numerical methods to find the roots of the polynomial and see which higher derivatives they satisfy. This was done for the polynomial

$$\begin{aligned} f(x) &= (x^2-2)^2(70x-99) \\ &= 70x^5 - 99x^4 - 280x^3 + 396x^2 + 280x - 396. \end{aligned}$$

The algorithm produced five "roots" for f:

$$\begin{aligned} x_1 &= 1.413665249 \\ x_2 &= -1.414218427 \\ x_3 &= 1.414803333 \\ x_4 &= -1.414208584 \\ x_5 &= 1.414244143 \end{aligned}$$

Additional possibilities for roots were obtained by applying the root finding algorithm to f' and f'', retaining values in the vicinity of the x_i above. We obtained

$$\begin{aligned} x_1' &= 1.414261664 \\ x_2' &= 1.414213562 \\ x_3' &= -1.414213562 \end{aligned}$$

using f' and

$$x_1'' = 1.414237614$$

using f''.

Here is a table of approximate values of f, f', and f'' for a few of these

x	$f(x)$	$f'(x)$	$f''(x)$
x_1	0	$5.5(-4)$	-1.9
x_2'	0	0	$-8(-2)$
x_1''	0	$-9(-7)$	$3.4(-6)$

(The numbers enclosed in parentheses are exponents.)

It is clearly quite difficult to determine from this whether f has a triple root near 1.4142, three closely spaced simple roots, or (correctly) a double root and nearby simple root.

1.2. The important fact to be understood is that a machine does not do arithmetic with real numbers. The machine represents nonzero numbers in the form

$$\pm .a_1a_2a_3 \ldots a_t \cdot b^e, \tag{1}$$

where b is usually 2 or 10 and the a_i are, correspondingly, binary or decimal digits. Most computers use the binary representation.

The number of digits, t, and the range of the exponent, e, depend upon the machine. Numbers of the form (1) together with 0 will be called (t-digit) machine numbers.

When an arithmetic operation is performed on two machine numbers, the exact result may not be a machine number (e.g., the product of two t-digit numbers will generally require $2t$ digits to be represented exactly.) In many language implementations the computer will round the result to the nearest machine number.

The values obtained in the example are typical of what occurs as a result of rounding error. Any method for finding numerical values for the roots of a polynomial $f(x)$ depends upon information provided by computed values of $f(x)$. If x is sufficiently close to a root the error in computing $f(x)$ will be comparable in magnitude to $f(x)$ itself. When this occurs we cannot even be sure that the computed sign of $f(x)$ is correct. Since the magnitude of $f(x)$ decreases rapidly near roots of high multiplicity, we expect that the band in which the error is significant will be especially large near a multiple root. It is important to note that the problem does not stem from the particular root finding algorithm used, but rather from the inability of a machine using fixed precision floating-point arithmetic to compute accurately the value of $f(x)$.

A reader interested in learning more about the consequences of using floating-point arithmetic is directed to the excellent book of Wilkinson [**8**]. Wilkinson provides a detailed analysis of the effect of rounding error on standard computations.

2. "Exact" Arithmetic.

2.1 If we think of a computation taking a set of input numbers and producing a set of output numbers, the effect of roundoff error will generally be to produce as output the result which would have been obtained had the computation been performed exactly on slightly perturbed input. In algebraic geometry, we are usually interested in "exceptional" phenomena: singularities and the like. Small perturbations in coeffficients of polynomials will change entirely the behavior of the objects under study. We must expect, therefore, to have difficulty obtaining desired information if computations are subject to roundoff errors.

It is first of all necessary that the data for a computation be in a form which the machine can represent exactly (rather than approximately). Second, computations should be performed using an "exact arithmetic" so that no roundoff errors can occur. Geometers have, in this respect, one very decided advantage: we are usually interested in examining examples to gain insight. Any two examples having the same general behavior are equivalent from this point of view. We can therefore often choose our examples to be compatible with the demands of machine computation.

[Continued on p. 45.]

2.2. In this section we will assume that our polynomials have integer coefficients. We will allow the operations of addition, subtraction, and multiplication. Division will only be performed in the "exact" sense of the Euclidean algorithm. We formulate the "root multiplicity" problem so that it is capable of solution in this context.

If $f(x)$ is in $\mathbb{Z}[x]$ it may be factored in the form

$$f(x) = c\pi_1^{\nu_1} \cdots \pi_m^{\nu_m} \tag{2}$$

where the π_i are distinct irreducible primitive polynomials.

DEFINITION 2.1. If $f(x)$ is factored as in (2), let $f_s(x)$ be the polynomial defined by $f_s(x) = \prod_s \pi_i$ where $\prod_s$ indicates a product over $\{i \mid \nu_i = s\}$.

With this definition, we have

$$f = c\prod (f_s)^s \tag{3}$$

where the f_s are primitive polynomials having only simple roots and where f_i and f_j are relatively prime if $i \neq j$.

It is clear that the roots of multiplicity s of $f(x)$ are precisely the roots of $f_s(x)$. If the factorization (3) is obtained we will be able to say with precision how many roots f has of various multiplicities. Numerical methods can be used to provide approximate locations of the roots if desired. (Since the f_s have simple roots, numerical methods will generally be more successful when applied to the f_s rather than to f.)

The task of factoring a polynomial with integer coefficients into irreducible factors, like the task of factoring an integer into primes, can be accomplished constructively [7] but is generally quite time-consuming from a computational point of view. It is somewhat surprising that the factorization (3) can be obtained with far less labor.

DEFINITION 2.2. Given $f \in \mathbb{Z}[x]$ define $g_i \in \mathbb{Z}[x]$ by
(i) $g_0 = f$
(ii) $g_{i+1} = \mathrm{GCD}(g_i, g_i')$.

DEFINITION 2.3. With the g_i as above, define h_i, f_i $(i = 1,2,3,\dots)$ by
(i) $g_{i-1} = g_i h_i$
(ii) $h_i = h_{i+1} f_i$ (it will be established that $h_{i+1} \mid h_i$).

THEOREM 2.4. *The f_i which appear in Definition* 2.3 *coincide with those which occur in factorization* (3).

Proof. Let $f = c\prod (f_s)^s$ as in (3). We will show that $g_i = c\prod_{\nu > i} (f_\nu)^{\nu - i}$. This is clearly true for g_0. In view of the form of the assertion it suffices to establish it for $i = 1$. We have

$$f' = c\left(f_2 f_3^2 \cdots f_k^{k-1}\right)\left[(f_1' \cdots f_k) + \cdots + (k f_1 \cdots f_k')\right] \tag{4}$$

We obtain our result by observing that the polynomial in square brackets is relatively prime to f. Indeed, if π is an irreducible primitive polynomial which divides f it must divide precisely one of the f_i. It will then divide all terms in the square brackets except for $if_1 \cdots f_i' \cdots f_k$. It cannot divide f_j for $j \neq i$ since the f_j are relatively prime. It cannot divide f_i' since f_i has no multiple roots.

It follows from this, then, that $h_i = \prod_{\nu \geqslant i} (f_\nu)$. We see, therefore, that $h_{i+1} \mid h_i$ and that the f_i of Definition 2.3 coincide with the f_i in factorization (3).

2.3. The problem of determining root multiplicities has now been reduced to the problem of computing the GCD of two polynomials with integer coefficients. $\mathbb{Z}[x]$ is not a euclidean domain but $\mathbb{Q}[x]$ is. It is possible, then, to implement the Euclidean Algorithm by using exact arithmetic in the field $\mathbb{Q}$. An extension of the Euclidean Algorithm which allows us to stay entirely within $\mathbb{Z}$ is found in many books on algebra (see, for example, [**3**, Theorem 2.14]). Using this algorithm we

have a procedure, performable entirely within $\mathbb{Z}[x]$, which solves the root multiplicity problem. Only exact operations on integers are required for carrying out the process.

If the process is implemented on a machine which carries out arithmetic with a fixed number of digits, the only source of error would be the appearance of integers whose lengths exceed the digit-carrying capacity of the machine. It is therefore important to analyze how integers grow in the execution of the algorithm. If S is an upper bound for the absolute value of the coefficients of a polynomial $f(x)$, we can estimate by recurrence the size of coefficients which could appear in the computation of GCD(f, f'). The analysis provides discouraging results: if $f(x)$ has degree n, then we can expect the appearance of coefficients of magnitude S^{e_n} where

$$e_1 = 1, \quad e_2 = 3, \quad e_{m+1} = 2e_m + e_{m-1}. \tag{5}$$

An indication of the growth of the e_n is seen in the table:

n	1	2	3	4	5	6	7	10	20
e_n	1	3	7	17	41	99	239	3363	22,619,537

Remember that the e_n are EXPONENTS!!!

Example. The algorithm was performed to find GCD(f, g) with

$$f = 11x^3 - 7x^2 + 5x - 3$$
$$g = 19x^2 + 17x + 13.$$

The result was GCD$(f, g) = 1$. The largest number which appeared in the process was 266,536,192.

The situation which arises here seems to be quite common in computations involving polynomials with integer coefficients. The input and output for an algorithm may be integers of relatively small size; yet extremely large integers occur in the midst of the algorithm.

The first thing that strikes someone who decides to do machine computation in algebra or algebraic geometry is that a great deal of modern mathematics is not even presented "constructively." The problem that we now face, however, illustrates the difference between "constructive" and "computational" mathematics: we may have at hand a procedure for theoretically computing some object of interest to us, but there are barriers to actually implementing the procedure. There is a difference between computing something "in principle" and computing something "in fact"!

The problem of efficiently computing the GCD of two polynomials with integer coefficients has received a great deal of attention [**1**], [**2**], [**4**]. Knuth [**4**, p. 369] describes more sophisticated algorithms which provide more modest growth rates for the coefficients. Computer systems designed for algebraic computation generally allow "infinite precision" integer arithmetic and these algorithms can be implemented on systems of this type. Machines which perform fixed precision arithmetic often allow an accuracy of only between 6 and 16 decimal digits before roundoff occurs. The growth rate for these revised algorithms is still too great to make them useful for fixed precision machines.

3. Modular Algorithms. The fact that the actual output from the GCD algorithm will be small if the input numbers are small gives rise to the hope that the difficulties outlined in the previous section can be surmounted by a change in algorithm. A technique for doing this was developed by G. E. Collins and W. S. Brown (see [**1**], [**4**]). The idea is to reduce the coefficients of polynomials modulo a prime p. The GCD computations are then carried out mod p. If $p < b^{t/2}$ (where $b = 2$ or 10 depending on whether the machine uses binary or decimal arithmetic) all computations mod p can be carried out on a machine using t-digit arithmetic. This is done for several primes p and the GCD over the integers is obtained by use of the Chinese Remainder Theorem.

If we employ the algorithms discussed in the previous section which use only operations within the integers, then for all but a finite number of primes the calculations mod p run exactly parallel

to the corresponding calculations over the integers. For some p the argument given in the proof of Theorem 2.4 will not apply. An illustration of what can occur is given by the following example.

Example. Let $f(x) = x^6 + x^3 + 1 \in \mathbb{Z}_3[x]$. In this case we have $f'(x) = 0$, and so the sequence of polynomials in Definition 2.2 is $g_i = f$ for all i. As a result we find $h_i = 1$ for all i. In fact, though, $f = (x^2 + x + 1)^3 = (x - 1)^6$. The argument breaks down at the point where we assert that $(f_1' \cdots f_k) + \cdots + (kf_1 \cdots f_k')$ is relatively prime to f. In this example, $f_1 = \cdots = f_5 = 1$ and $f_6 = x - 1$. The expression reduces to $6f_6'$ which $= 0$ in $\mathbb{Z}_3[x]$. The failure occurs because the coefficient 6 vanishes. If $p > \deg f$, this problem cannot occur. The rest of the argument is valid mod p.

Thus, if we choose $p > \deg f$, the procedure will give a factorization mod p analogous to (3).

Now let $f \in \mathbb{Z}[x]$. Factor f as in (3) and let

$$\bar{f} = \bar{c}\Pi(\bar{f}_s)^s \tag{6}$$

where $\bar{}$ indicates reduction modulo p ($p > \deg f$).

THEOREM 3.1. *There are only a finite number of exceptional primes for which* (6) *is not the result obtained by the* mod p *algorithm.*

Proof. Since we have assumed that $p > \deg f$, the mod p algorithm will produce a factorization which can only fail to coincide (up to associates) with (6) if the $\bar{f}_i$ are not pairwise relatively prime or if some $\bar{f}_i$ has multiple roots. Since f_i and f_j are relatively prime in $\mathbb{Z}[x]$, $\bar{f}_i$ and $\bar{f}_j$ can only have a common factor mod p if p divides the resultant $R(f_i, f_j)$. Similarly, since $f_i \in \mathbb{Z}[x]$ has no multiple roots, $\bar{f}_i$ can only have a multiple root if p divides $R(f_i, f_i')$.

After the factorization mod p is determined for several primes p we get the factorization (3) by use of the Chinese Remainder Theorem. A check that an exceptional prime was not used is that the f_i obtained are indeed factors of f (this also provides a way of checking whether sufficiently many primes have been used).

Example. Let

$f(x) = 128625x^7 - 236425x^6 + 173110x^5 - 66914x^4 + 14929x^3 - 1937x^2 + 136x - 4.$

Let

$$p_1 = 99991 \quad \text{and} \quad p_2 = 99989.$$

For $p = 99991$ we obtain:

$$f_1^{(1)} = 37141x^2 - 11565x + 2313$$
$$f_2^{(1)} = -5567x - 42058$$
$$f_3^{(1)} = -242x - 39948.$$

For $p = 99989$ we obtain:

$$f_1^{(2)} = 5509x^2 - 33814x - 13235$$
$$f_2^{(2)} = 42436x + 22506$$
$$f_3^{(2)} = 9738x + 38048.$$

Provided the p_i are not exceptional, these are essentially the results we would obtain by reducing mod p_i the f_j obtained by applying the algorithm in $\mathbb{Z}$. "Essentially" means up to a constant multiple. We determine constants c_1, c_2 so that the leading coefficients of $c_i f_j^{(i)}$ ($i = 1, 2$) are the reductions mod p of the leading coefficient of f. The Chinese Remainder Theorem is applied to $c_1 f_j^{(1)}$ and $c_2 f_j^{(2)}$ and the resulting polynomials are made primitive. We find

$$f_1 = 21x^2 - 20x + 4$$
$$f_2 = 7x - 1$$
$$f_3 = 5x - 1.$$

It is verified that these are indeed factors of f.

The original polynomial has two simple roots, a double root, and a triple root. It might be of interest to note that the exceptional primes are 2, 3, 5, 7, and 11.

4. Explicit Algorithms. The key to successful use of a computer is to ensure that correct algorithms are applied to correct data. In this section we will describe explicitly some of the algorithms to be used by the program and outline proofs of their correctness (within their operating limits). We will need:

1. algorithms for arithmetic modulo p
2. an algorithm for the GCD of polynomials mod p
3. an algorithm for the Chinese Remainder Theorem.

We will be implementing these algorithms on a machine which performs arithmetic with a fixed number, t, of b-ary digits.

4.1 The operations of addition, subtraction, and multiplication are performed mod p by performing them in $\mathbb{Z}$ on representatives and reducing the result mod p. If $x, y \in \mathbb{Z}$ satisfy $|x|, |y| < p$ then correct results will be obtained provided $p < b^{t/2}$. Division is performed as follows:

Let x represent a non-zero class mod p. Obtain integers A and B so that $Ax + Bp = 1$. The class of x^{-1} mod p is represented by A.

Algorithm A (Extended GCD Algorithm for Integers):

Input: Nonzero integers x and y

Output: $d = \mathrm{GCD}(x, y)$ and A, B so that $d = Ax + By$

1 $r_0 \leftarrow x, \quad A_0 \leftarrow 1, \quad B_0 \leftarrow 0,$
$r_1 \leftarrow y, \quad A_1 \leftarrow 0, \quad B_1 \leftarrow 1, \quad n \leftarrow 2$

2 $r_{n-2} = q_n r_{n-1} + r_n$
with $0 \leqslant r_n < r_{n-1}$ (Euclidean Algorithm)

3 If $r_n = 0$, then go to step 6

4 $A_n \leftarrow A_{n-2} - q_n A_{n-1},$
$B_n \leftarrow B_{n-2} - q_n B_{n-1}, \quad n \leftarrow n + 1$

5 go to step 2

6 $d \leftarrow r_{n-1}, \quad A \leftarrow A_{n-1}, B \leftarrow B_{n-1},$ stop

THEOREM 4.1. *If $|x|, |y| < b^{t/2}$, Algorithm A will produce correct results on a machine using t-digit, b-ary arithmetic.*

REMARK. When we implement the algorithm we will be using $y = p$ and will only compute A.

The proof will be given in the case $y > 0$ and will show that the A_i are computed correctly.

Proof.

$$r_{i-1} = r_i q_{i+1} + r_{i+1}, \tag{1}$$

$$\begin{aligned} A_{i+1} &= A_{i-1} - A_i q_{i+1} \\ B_{i+1} &= B_{i-1} - B_i q_{i+1} \end{aligned} \tag{2}$$

Using (1), (2), and the initial values of the A_i, B_i, and r_i we prove by induction that

$$A_i x + B_i y = r_i \tag{3}$$

and that

$$A_{i+1}B_i - A_i B_{i+1} = (-1)^i. \tag{4}$$

LEMMA 4.2. (i) $1 = A_2 < A_4 < \cdots$.
(ii) $-1 \geqslant A_3 > A_5 > \cdots$.
(iii) $|A_i| < |A_{i+1}|$ *for* $i = 2, 3, \ldots, n$
(iv) $A_n = |y|/|d|$

Proof of Lemma 4.2. From (2) and the initial values $A_0 = 1$, $A_1 = 0$ we find $A_2 = 1$, $A_3 = -q_3$. We have assumed that $y = r_1 > 0$ so that $q_i > 0$ for $i = 3, 4, \ldots$. Thus $-A_i q_{i+1} = A_{i+1} - A_{i-1}$ is < 0 for i even and > 0 for i odd. This establishes (i) and (ii).

We also have

$$\frac{A_{i+1}}{A_i} = \frac{A_{i-1}}{A_i} - q_{i+1}.$$

Since the A_i alternate in sign we obtain

$$\frac{|A_{i+1}|}{|A_i|} = q_{i+1} + \frac{|A_{i-1}|}{|A_i|} > q_{i+1} \geqslant 1,$$

which proves (iii).

From (3) we see that, in particular, $A_n x + B_n y = 0$. (4) shows that A_i and B_i are relatively prime, so that $|A_n| = |y|/|d|$. This completes the proof of the Lemma.

Lemma 4.2 shows that $|A_i| < |y| < b^{t/2}$. In step 4 both the multiplication and subtraction are correctly performed since we also have $|q_i| \leqslant |x| < b^{t/2}$.

4.2. Algorithm B (GCD of Polynomials Mod p):

Input: polynomials $f = a_n x^n + \cdots + a_0$

$g = b_m x^m + \cdots + b_0$ in $\mathbb{Z}_p[x]$

Output: $R = \mathrm{GCD}(f, g)$

1 If $n < m$ then $f \leftrightarrow g$

2 If $g = 0$ then $R \leftarrow f$, stop

3 $h \leftarrow \mathrm{Red}_p(b_m f - a_n x^{n-m} g)$

4 $f \leftarrow g$, $g \leftarrow h$, go to step 1

We will assume that an element $\bar{x}$ of $\mathbb{Z}_p$ is represented by an integer x which satisfies $-\frac{1}{2}p < x \leqslant \frac{1}{2}p$. $\mathrm{Red}_p(h)$ denotes the operation which reduces the coefficients of h mod p (using

representatives in the range indicated). The key to the correctness of step 3 on a t-digit machine is that $AB + CD$ can be correctly computed provided that $|A|, |B|, |C|, |D| \leq \frac{1}{2}p < \frac{1}{2}b^{t/2}$. We leave to the reader the rest of the verification that, with $p < b^{t/2}$, the algorithm will produce correct results.

4.2. Algorithm C (Chinese Remainder Theorem):

Input m_i, a_i $(i = 1, 2)$ integers with
$m_1, m_2 > 0$ relatively prime
Output: x satisfying $x \equiv a_i \bmod m_i$ $(i = 1, 2)$

1 $a_i \leftarrow \mathrm{cl}_i(a_i)$
2 if $m_1 < m_2$ then $m_1 \leftrightarrow m_2, a_1 \leftrightarrow a_2$
3 $Am_1 + Bm_2 = 1$ (Algorithm A)
4 $k \leftarrow a_2 - a_1$, $\lambda = \mathrm{cl}_2(Ak)$
5 $x \leftarrow a_1 + \lambda m_1$, stop

In this algorithm $\mathrm{cl}_i(x)$ is used to denote the representative of the class of x modulo m_i which satisfies $-\frac{1}{2}m_i < x \leq \frac{1}{2}m_i$.

THEOREM 4.3. *Algorithm C produces correct results if*

$$m_1 m_2 < b^t.$$

Proof. The correctness of Algorithm A has already been established. We obtain an integer A satisfying $|A| \leq \frac{1}{2}m_2$. We have $|k| \leq \max(m_1, m_2) = m_1$. If $m_1 m_2 < b^t$ the product Ak is correctly computed and $|\lambda| \leq \frac{1}{2}m_2$. We find that $|\lambda m_1| = |\lambda||m_1| \leq \frac{1}{2}m_1 m_2$ and, since $|a_1| \leq \frac{1}{2}m_1 \leq \frac{1}{2}m_1 m_2$, x is correctly computed in step 5.

It is obvious that $x \equiv a_1 \bmod m_1$. We have

$$x \equiv a_1 + Akm_1 \equiv a_1 + k \equiv a_1 + a_2 - a_1 \equiv a_2 \bmod m_2.$$

COMMENT: In spite of the brevity of this section, much of the time involved in this project was devoted to these matters. Since a computer is generally used to provide results which would be too time consuming to obtain by hand, the user's only insurance against incorrect results is:

1. Correct data has been supplied.
2. Correct algorithms have been applied.
3. Limitations on the validity of the algorithms are understood and adhered to.

5. Implementation on Real Machines. Implementation introduces a new set of concerns: the choice of language, the representation of data, the management of storage, the type of user interaction. This article includes a computer program to serve as an example of implementation on a real machine. The program is written in BASIC—the most common language found on microcomputers. Specifically, the program is written in Level II BASIC for the TRS-80. We have avoided heavy use of machine-specific commands so that readers whose computers use another dialect of BASIC can modify the program to run on their machines. (See pp. 51–53.)

5.1. *Program Organization.* The program is a collection of subroutines followed by a main program. Each subroutine is prefaced by a remark which describes its function and names the parameters which must be passed to it by a calling routine. Subroutines to perform arithmetic on

```
1 REM
      ROOT MULTIPLICITY PROGRAM       VERSION 2.3
        J. WAVRIK      JULY 3, 1980

10 DEFINT D,J,K
20 DIM P(53,10),D(53),MD(53)
30 GOTO 5000
98 '
      ++++++++++++++++++++++++++++
99 '  SUBROUTINE TO COPY P(JS,.) TO P(JT,.)
      PARAMETERS: JS,JT

100 D(JT)=D(JS) : MD(JT)=MD(JS)
110 FOR J=0 TO D(JS) : P(JT,J) = P(JS,J) : NEXT J
120 RETURN
198 '
      ++++++++++++++++++++++++++++
199 '  SUBROUTINE FOR GCD ALGORITHM STEP 3
      PARAMETERS: JF, JG, JH

200 M=D(JG) : N=D(JF) : D(JH)=N
210 A=P(JF,N) : B=P(JG,M)
220 FOR J=0 TO N-M-1
230 P(JH,J)=B * P(JF,J)
240 NEXT J
250 FOR J=N-M TO N
260 P(JH,J) = B * P(JF,J) - A * P(JG,J+M-N)
270 NEXT J
280 JR = JH
298 '
      ++++++++++++++++++++++++++++
299 '  SUBROUTINE TO REDUCE P(JR,.) MOD P
      PARAMETER: JR

300 J1=0 : MD(JR)=P
310 FOR J = 0 TO D(JR)
320 X = P(JR,J)
330 X = X - P * INT(X/P)
340 IF 2 * X > P THEN X = X - P
350 P(JR,J) = X
360 IF X <> 0 THEN J1=J
370 NEXT J
380 D(JR)=J1 : RETURN
398 '
      ++++++++++++++++++++++++++++
399 '  SUBROUTINE FOR H = GCD(F,G)  (MOD P)
      PARAMETERS KF, KG, KH

400 JF=51 : JG=52
410 JT=JF : JS=KF : GOSUB 100
420 JT=JG : JS=KG : GOSUB 100
430 IF D(JF) < D(JG) THEN JS=JF:JF=JG JG=JS    'STEP 1
440 IF P(JG,D(JG))=0 THEN JR=JF GOTO 480       'STEP 2
450 JH = 156 - JF - JG         'MAKE JH THE "OTHER" INDEX
460 GOSUB 200                  ' STEP 3
470 JF=JG . JG=JH : GOTO 430  'STEP 4
480 JT=KH : JS=JR : GOTO 100
498 '
      ++++++++++++++++++++++++++++
499 '  SUBROUTINE FOR G = GCD(X,Y)
      PARAMETERS: X,Y (INTEGERS)

500 IF Y=0 THEN G=X : RETURN
510 R0 = ABS(X) : R1 = ABS(Y)
520 Q = INT(R0/R1) : R = R0 - R1 * Q
530 IF R <> 0 THEN R0 = R1 : R1 = R : GOTO 520
540 G = R1 : RETURN
598 '
      ++++++++++++++++++++++++++++
599 '  SUBROUTINE TO COMPUTE Y = INVERSE OF X MOD P

600 A0=1 : A1=0 : R0=X : R1=P
610 Q = INT(R0/R1) : R = R0 - R1 * Q
620 IF R = 0 THEN 650
630 A = A0 - Q * A1 : A0=A1 : A1=A
640 R0=R1 : R1=R : GOTO 610
650 Y=A1 : IF Y < 0 THEN Y = Y + P
660 IF 2*Y > P THEN Y = Y - P
670 RETURN
698 '
      ++++++++++++++++++++++++++++
699 '  SUBROUTINE TO MULTIPLY P(JR,.) BY C
      PARAMETERS: JR,C

700 FOR J=0 TO D(JR) : P(JR,J) = P(JR,J) * C : NEXT J
710 GOTO 300
798 '
      ++++++++++++++++++++++++++++
799 '  SUBROUTINE FOR A = BQ+R MOD P  (WITH B MONIC)
      PARAMETERS: JA,JB,JQ,JR

800 D(JQ)=0 : P(JQ,0) =0
810 IF D(JA) < D(JB) THEN JR=JA : RETURN
820 D(JQ) = D(JA) - D(JB)
830 FOR J = D(JQ) TO 0 STEP -1
840 P(JQ,J) = P(JA,D(JA))
850 FOR J1 = 0 TO D(JB)
860 PX = P(JA,J1+J) - P(JQ,J) * P(JB,J1)
870 PX = PX - P * INT(PX/P)
880 IF 2 * PX > P THEN PX = PX - P
890 P(JA,J1+J) = PX
900 NEXT J1
910 D(JA) = D(JA)-1 : NEXT J
920 IF D(JA) < 0 THEN D(JA) = 0
930 JR=JA : RETURN
998 '
      ++++++++++++++++++++++++++++
999 '  SUBROUTINE FOR A = BQ + R  MOD P
      PARAMETERS: KA,KB,KQ,KR

1000 JA=51 : JB=52 : JQ=53
1010 JT=JA : JS=KA : GOSUB 100
1020 JT=JB : JS=KB : GOSUB 100
1030 X = P(JB,D(JB)) : GOSUB 600
1040 C=Y : JR=JB : GOSUB 700
1050 GOSUB 800
1060 JT=KR : JS=JR : GOSUB 100
1070 JT=KQ : JS=JQ : GOSUB 100
1080 JR=KQ : C=Y : GOSUB 700
1090 JR=KR : GOTO 300
1098 '
      ++++++++++++++++++++++++++++
1099 ' OUTPUT DISPLAY OF CURRENT PAGE

1100 CLS
1110 FOR J1=0 TO 9: J2 = 10 * PG% + J1
1120 PRINT J2"." TAB(10);
1130 FOR J = D(J2) TO 0 STEP -1
1140 PRINT P(J2,J); : NEXT J
1150 PRINT : NEXT J1
1160 PRINT : RETURN
1198 '
      ++++++++++++++++++++++++++++
1199 ' OUTPUT CURRENT PAGE WITH SELECTED TITLES
      PARAMETERS: T$,KS,JM

1200 GOSUB 1100
1210 JX = KS - 10 * INT(KS/10)
1220 FOR JJ = 1 TO JM - KS
1230 TP$ = T$ + CHR$(48+JJ)
1240 PRINT @ 64 * (JX+JJ) + 6, TP$;
1250 NEXT JJ
1260 RETURN
1298 '
      ++++++++++++++++++++++++++++
1299 ' SUBROUTINE TO SET P(JD,.) = DERIVATIVE OF P(JO,.)
      PARAMETERS: JD, JO

1300 D(JD) = D(JO) - 1
1310 IF D(JD) < 0 THEN D(JD)=0 : P(JD,0)=0 : RETURN
1320 FOR J = 0 TO D(JD)
1330 P(JD,J) = P(JO,J+1) * (J+1)
1340 NEXT J
1350 JR=JD : GOTO 300
```

```
1398 '
     ++++++++++++++++++++++++++++++++++
1399 ' ALGORITHM FOR FACTORIZATION MOD P
     PARAMETER: KS   (LOCATION FOR F)

1400 JM=KS
1410 JO=JM : JD=JM+1 : GOSUB 1300
1420 KF=JO : KG=JD : KH=JD : GOSUB 400
1430 JM = JM + 1
1440 IF D(JM) > 0 AND D(JM) < D(JM-1) THEN 1410
1450 T$="G" : GOSUB 1200
1460 FOR JI = JM TO KS+1 STEP -1
1470 KA = JI-1 : KB=JI : KQ=JI : KR=50
1480 GOSUB 1000
1490 NEXT JI
1500 T$="H" : GOSUB 1200
1510 IF JM = KS+1 THEN 1570
1520 FOR JI = KS+1 TO JM-1
1530 KA=JI : KB=JI+1 : KQ=JI : KR=50
1540 GOSUB 1000
1550 NEXT JI
1560 T$="F" : GOSUB 1200
1570 FOR JI = KS+1 TO JM
1580 X = P(JI,D(JI)) : GOSUB 600
1590 C = Y * P(KS,D(KS)) : C = C - P * INT(C/P)
1600 IF 2 * C > P THEN C = C - P
1610 JR = JI : GOSUB 700
1620 NEXT JI
1630 T$="F" : GOSUB 1200
1640 PRINT @ 704, ""; : RETURN
1698 '
     ++++++++++++++++++++++++++++++++++
1699 ' SUBROUTINE FOR CHINESE REMAINDER THEOREM
     PARAMETERS: K1,K2,KR

1700 IF D(K1) <> D(K2) THEN 1940
1710 X=MD(K1) : Y=MD(K2) : GOSUB 500
1720 IF G <> 1 THEN 1950
1730 JT=51 : JS=K1 : GOSUB 100
1740 JT=52 : JS=K2 : GOSUB 100
1750 J1=51 : J2=52 : J3=53
1760 IF MD(J1) < MD(J2) THEN J1=52 : J2=51
1770 X=MD(J1) : P=MD(J2) : GOSUB 600
1780 D(J3)=D(J1) : MD(J3) = MD(J1) * MD(J2)
1790 FOR J = 0 TO D(J3)
1800 X = Y * (P(J2,J) - P(J1,J))
1810 X = X - P * INT(X/P) : IF 2 * X > P THEN X = X - P
1820 P(J3,J) = P(J1,J) + X * MD(J1)
1830 NEXT J
1840 P = MD(J3)
1850 JT=KR : JS=J3 : GOTO 100
1860 '  ERROR MESSAGE
1870 PRINT @ 704,CHR$(31);
1880 FOR J = 1 TO 4
1890 PRINT @ 710,EM$;
1900 FOR J1 = 1 TO 300 : NEXT J1
1910 PRINT @ 710, CHR$(30);
1920 FOR J1 = 1 TO 300 : NEXT J1
1930 NEXT J : RETURN
1940 EM$="****  INCOMPATIBLE DEGREES  ****" : GOTO 1870
1950 EM$="****  MODULI NOT RELATIVELY PRIME  ****" : GOTO 1870
1998 '
     ++++++++++++++++++++++++++++++++++
1999 ' SUBROUTINE FOR MAKING P(KP,.) PRIMITIVE
     PARAMETER: KP

2000 IF D(KP) > 0 THEN 2030
2010 IF P(KP,0) <> 0 THEN P(KP,0)=1
2020 RETURN
2030 X = P(KP,0)
2040 FOR J=1 TO D(KP)
2050 Y = P(KP,J) : GOSUB 500
2060 X=G : NEXT J
2070 FOR J = 0 TO D(KP)
2080 P(KP,J) = P(KP,J)/G
2090 NEXT J : RETURN
2098 '
     ++++++++++++++++++++++++++++++++++
2099 ' SUBROUTINE TO ATTEMPT A = BQ + R OVER INTEGERS
     PARAMETERS: KA,KB,KQ,KR

2100 JA=51 : JB=52 : JQ=53
2110 JT=JA : JS=KA : GOSUB 100
2120 JT=JB : JS=KB : GOSUB 100
2130 D(JQ)=0 : P(JQ,0)=0
2140 IF D(JA) < D(JB) THEN 2240
2150 D(JQ) = D(JA) - D(JB)
2160 FOR J = D(JQ) TO 0 STEP -1
2170 P(JQ,J) = P(JA,D(JA))/P(JB,D(JB))
2180 IF P(JQ,J) <> INT(P(JQ,J)) THEN 2270
2190 FOR J1 = 0 TO D(JB)
2200 P(JA,J1+J) = P(JA,J1+J) - P(JQ,J) * P(JB,J1)
2210 NEXT J1
```

```
2220 D(JA) = D(JA)-1 : NEXT J
2230 IF D(JA) < 0 THEN D(JA) = 0
2240 JT=KR : JS=JA : GOSUB 100
2250 JT=KQ : JS=JQ : GOSUB 100
2260 RETURN
2270 EM$="**** NOT POSSIBLE OVER THE INTEGERS ****" : GOTO 1870
2298 '
     +++++++++++++++++++++++++++++++++++++++
2299 ' SUBROUTINE FOR INPUT OF P(KI,.)
     PARAMETER: KI

2300 INPUT "STORAGE LOCATION"; KI
2310 INPUT "DEGREE OF POLYNOMIAL"; D(KI)
2320 FOR J = D(KI) TO 0 STEP -1
2330 PRINT J "-TH COEFFICIENT"; : INPUT P(KI,J)
2340 NEXT J : RETURN

4999 '

     +++++ MAIN PROGRAM +++++

5000 PRINT "MENU","MODULUS = "P
5010 PRINT"I=INPUT, C=COPY, M=MODULUS INPUT, R=REDUCE, E=EUCL ALG MOD P"
5020 PRINT "Z=INTEGER EUCL ALG, D=DERIV, * = POLY * C, F=FACTOR,Q= CRT"
5030 PRINT "P=MAKE PRIM, Y= INV OF X, G= GCD"

5040 M$ = CHR$(10) + CHR$(91) + "ICMREZD*QPGYF"
5050 A$ = INKEY$ : IF A$ = "" THEN 5050
5060 FOR A% = 1 TO 15
5070 IF A$ = MID$(M$,A%,1) THEN 5100
5080 NEXT A% : GOTO 5050
5100 PRINT @ 704, CHR$(31);
5110 ON A% GOSUB 5200,5250,5300,5350,5400,5450,5500,5550,5600,
          5650,5700,5750,5800,5850,5900
5120 IF A% < 14 THEN GOSUB 1100
5130 GOTO 5000

5199 ' ***** TURN TO NEXT PAGE ****
5200 PG% = PG% + 1 : IF PG% > 4 THEN PG% = 0
5210 RETURN
5249 ' ***** TURN TO PREVIOUS PAGE ****
5250 PG% = PG% - 1 : IF PG% < 0 THEN PG% = 4
5260 RETURN
5300 TT$ = "**** INPUT POLYNOMIAL ****" : GOSUB 6000
5310 GOTO 2300
5350 TT$ = "**** COPY A POLYNOMIAL ****" : GOSUB 6000
5360 INPUT "SOURCE LOCATION, TARGET LOCATION"; JS,JT
5370 GOTO 100
5400 TT$ = "**** INPUT MODULUS ****" : GOSUB 6000
5410 INPUT "MODULUS"; P : RETURN
5450 TT$ = "**** REDUCE POLYNOMIAL MOD P ****" : GOSUB 6000
5460 INPUT "LOCATION # FOR POLYNOMIAL"; JR
5470 GOTO 300
5500 TT$ = "**** EUCLIDEAN ALGORITHM MOD P ****" : GOSUB 6000
5510 PRINT TAB(10) "A = BQ + R  MOD P"
5520 INPUT "LOCATIONS FOR A,B,Q,R"; KA,KB,KQ,KR
5530 GOTO 1000
5550 TT$ = "**** EUCLIDEAN ALGORITHM OVER INTEGERS ****" : GOSUB 6000
5560 PRINT TAB(6) "GIVEN A,B WILL FIND Q,R IF POSSIBLE (A = BQ + R)"
5570 INPUT "LOCATIONS FOR A,B,Q,R", KA,KB,KQ,KR
5580 GOTO 2100
5600 TT$ = "**** DIFFFRENTIATE POLYNOMIAL MOD P ****" : GOSUB 6000
5610 INPUT "LOCATION FOR POLY, DERIVATIVE"; JO,JD
5620 GOTO 1300
5650 TT$ = "**** MULTIPLY POLYNOMIAL BY C ****" : GOSUB 6000
5660 INPUT "LOCATION #, C"; JR,C
5670 GOTO 700
5700 TT$ = "**** CHINESE REMAINDER THEOREM ****" : GOSUB 6000
5710 INPUT "LOCATION FOR FIRST, SECOND, RESULT"; K1,K2,KR
5720 GOTO 1700
5750 TT$ = "**** MAKE POLYNOMIAL PRIMITIVE ****" : GOSUB 6000
5760 INPUT "LOCATION #"; KP
5770 GOTO 2000
5800 TT$ = "**** GCD(F,G) OF POLYNOMIALS MOD P ****" : GOSUB 6000
5810 INPUT "LOCATION FOR F,G,RESULT"; KF,KG,KH
5820 GOTO 400
5850 TT$ = "**** Y = INVERSE OF X MOD P ****" : GOSUB 6000
5860 INPUT "X"; X : GOSUB 600
5870 PRINT @ 640, "X = " X, "Y = " Y, "P = "P CHR$(31)
5880 RETURN
5900 TTS = "**** FACTORIZATION ALGORITHM ****" : GOSUB 6000
5910 INPUT "LOCATION FOR F"; KS
5920 GOTO 1400

6000 TB = 31 - INT(LEN(TTS)/2)
6010 PRINT TAB(TB) TTS
6020 PRINTTAB(5) "(TO RETURN TO MENU PRESS M...TO CONTINUE PRESS SPACE)"
6030 K$ = INKEY$ : IF K$ = "" THEN 6030
6040 IF K$ = " " THEN RETURN
6050 IF K$ <> "M" THEN 6030
6060 PRINT @ 704, CHR$(31); : GOTO 5000
```

polynomials are useful in a variety of programs. This type of organization permits the creation of a library of subroutines which can be used in other programs. The program constitutes a "tool-kit" of polynomial operations. The "tool-kit" is not complete: it contains only those operations needed to solve the root multiplicity problem. The program has been included with this article to serve a didactic purpose; it is intended to be read. Readers are encouraged to modify it and add to it.

5.2 *Representation of Data.* A polynomial is most naturally represented as an array of coefficients. $f = a_n x^n + \cdots + a_0$ could, for example, be represented as an array $F(\ )$ with $F(J) = a_J$. We have, however, chosen BASIC as the programming language. In BASIC all variables are global and subroutines must explicitly name their arguments. A variety of problems arise if we literally name polynomials using arrays $F(\ )$, $G(\ )$, etc. We may think of F and G as names for specific blocks of memory in the computer's store. Operations like step 1 in algorithm B (which interchanges f and g) would require the displacement of all the coefficients of f and g. Furthermore, if we were to create a subroutine which performs some desired operation on F and G and wish to apply it to polynomials named A and B, we would have to move the coefficients of A and B to the appropriate locations (saving and later restoring any data currently in those locations). Rather than swapping labels on blocks of memory we have to move the data itself. This excessive movement of data would diminish the efficiency of the program.

We use, instead, a block of memory configured as a double array $P(I, J)$. The first index is regarded as a pointer to a specific polynomial; the second index is the index of a coefficient in that polynomial. In this way subroutines can be created to use pointer values as arguments; interchange operations involve an exchange of pointers rather than movement of coefficients.

A descriptive nomenclature has been chosen for pointers: KF, KG, KH, etc., have been used for pointers which have a global meaning (and should not be changed by a subroutine) while JF, JG, JH, etc., have been used for pointers which have a temporary meaning. A systematic nomenclature of this sort helps reduce unexpected changes when the name of a variable in a subroutine conflicts with a name used in a calling routine. $D(I)$ represents the degree of the polynomial $P(I,.)$. When $P(I,.)$ results from an algorithm using arithmetic modulo p, the modulus is saved as $MD(I)$.

5.3. *Storage Management.* The method chosen for data representation embodies a storage management scheme. We have used a scheme which is wasteful of storage since a block of memory of fixed size is reserved for the maximum number of polynomials of the highest required degree (whether or not the storage is actually used). The advantage is the simplicity of the scheme and the ease with which data can be accessed and unused storage retrieved.

The program allots space for 54 polynomials of maximum degree 10. Some routines are destructive of data and so polynomials $P(I,.)$ for $I = 50$ to 53 constitute a "scratchpad." Destructive routines transfer data to the scratchpad when they start and to desired locations when they end. In this way a user is assured that the original data remains intact or else can elect to have it overwritten by results.

5.4. *User Interaction.* A determination of how a program will be used is an important part of program design. Microcomputers are limited in memory in comparison with larger computers, but they are usually single-user machines. Microcomputers are very well suited for use in a highly interactive environment in which the user supplies judgment and guidance. In this way the computer becomes an electronic adjunct to the mathematician, playing the role of ("intelligent") pencil and paper. The program provided in this article (see pp. 51–53) will, for example, compute the f_i in the factorization but will leave the user to decide how many primes to use and whether or not a prime is exceptional. It will apply the Chinese Remainder Theorem when requested to do so, but not automatically. The computer and user have a symbiotic relationship. The program itself

does not incorporate a great deal of "knowledge" but is rather designed to provide flexible manipulative facility.

The program shows the user a "page" of 10 polynomials. The "page" can be turned so that the user can see all 50 polynomials (10 at a time). The program is "menu-driven": the user is provided a list of operations which can be executed and must choose from the list. Once a choice is made, the program will prompt the user for additional input if required. When a command is selected from the menu an "abort" option allows the user to return to the menu rather than executing the command.

5.5. *Comments on Details*. When a polynomial is produced by some operation, we need to calculate its degree. The subroutine starting at line 300 combines the computation of the degree with the reduction of the polynomial mod p. This subroutine is called at the end of many of the other routines.

The program will make a few checks for potential errors (see lines 1700, 1720, 2180) but no check is made that p is prime and within the range ($p < b^{t/2}$) in which the algorithms operate correctly.

Input and output sections of a program tend to be more machine-specific than other parts. The PRINT @ command (e.g., line 1240) is used to place output at a specific location on the CRT screen. The INKEY$ command (e.g., line 5050) reads a character from the keyboard without requiring a carriage return. CHR$(10) and CHR$(91) (line 5040) are the characters returned from the keyboard by the downward and upward pointing arrow keys (the program uses these keys to turn "pages" back and forth). CHR$(31) is used on the TRS-80 to clear the video screen from the current cursor position to the end of the screen.

The values of b and t (which determine the range in which the algorithms are valid) are machine dependent. The algorithms were first developed on an HP-97 programmable calculator which uses a binary-coded decimal representation for numbers. In this case $b = 10$ and $t = 10$. The TRS-80 supports three numerical data types: integer, single precision, and double precision. Single precision uses a binary representation with an 8-bit exponent and a 24-bit mantissa. Double precision uses an 8-bit exponent and a 56-bit mantissa. On this machine, then, $b = 2$ and $t = 24$ or 56. The algorithms will work correctly if $p < 4096$ (for single precision) or $p <$ 268,435,456 (for double precision).

Double precision is obtained by adding the line

```
5 DEFDBL A-Z
```

to the program. The program currently uses single precision (by default).

6. Final Remarks.

6.1. The purpose of this article is to give readers some of the flavor of algebraic computation. A program is included, not to provide readers with a piece of ready-made software, but rather to give them ideas for developing their own program. Readers whose background is in pure mathematics will find the concerns of machine computation novel. An interest in machine computation presents, for a pure mathematician, entry into another field. The mathematics literature, however, is not organized in a way which easily permits one to answer the question "What is known about XYZ?" This article is a revision of an earlier version in which the author, in blissful ignorance, represented his rediscovery of the modular algorithm for GCDs. Professors Joseph Traub and Charles Sims were kind enough to comment on the earlier version and provide some access to the relevant literature. In this section we will provide a similar service for the reader.

A good starting point, for someone seeking entry to the subject of machine computation, is the series *The Art of Computer Programming*, by Donald E. Knuth. This is a proposed seven-volume

set, of which the first three volumes have appeared.

"Numerical analysis" is a branch of mathematics which deals with computation. Generally, however, the numerical analysis literature discusses the sort of computations which arise in science. The heading under which literature on algebraic computation is found is "Symbolic and Algebraic Manipulation." As we pointed out in Section 1, most computer users interact primarily with a language or system. There are several systems which have been developed specifically for symbolic and algebraic computations (e.g., ALTRAN, CAMAC, MACSYMA, REDUCE, SAC, SCRATCHPAD). Most of these require the resources of a large time-share computer system and are only implemented at a few sites. A microcomputer algebra system, muMATH, has been developed by Rich and Stoutemeyer [**6**]. A few of these systems are for special purposes (e.g., the study of groups). The general-purpose systems seem to concentrate on "high school algebra" through "college calculus": expansion and simplification of rational expressions, symbolic integration and differentiation, etc.

There are many types of problems and computations which could interest readers of this article. Rather than give a bibliography of articles we give a list of names of some of the workers in the field to assist in initiating a literature search. Apology is given to those whose names have been omitted—no insult is intended, the list is not exhaustive!

E. R. Berlekamp, G. Birkhoff, W. S. Brown, B. F. Caviness, G. E. Collins, J. H. Conway, J. Davenport, R. J. Fateman, A. Hearn, D. E. Knuth, D. Lazard, D. and E. Lehmer, W. A. Martin, J. Moses, J. Neubuser, E. W. Ng, V. Pless, C. C. Sims, B. Trager, J. F. Traub, D. Y. Y. Yun, H. Zassenhaus, H. Zimmer.

The Special Interest Group on Symbolic & Algebraic Manipulation of the Association for Computing Machinery (ACM) publishes a quarterly bulletin. The group is known as SIGSAM. (The abundant use of acronyms is a feature that those approaching computation from the outside might find rather amusing. The names are apparently chosen by a group of people who did excessive amounts of assembly language programming in their youth—this has affected their whole outlook on life!) There are periodic conferences, symposia, etc., on symbolic and algebraic manipulation. The proceedings of these conferences provide both surveys of current developments and further references.

Papers on algebraic computation are currently published primarily in journals devoted to computer science. At times, an author of a paper in a mathematics journal will mention that a computer has been used to assist in the research. The details of the algorithms rarely seem to appear, however. I have no idea whether most mathematics journals have a policy preventing the publication of computer algorithms.

A few textbooks on algebraic computation are currently in preliminary form and are soon to be published.

References

1. W. S. Brown, On Euclid's algorithm and the computation of polynomial greatest common divisors, J. ACM, 18 (Oct. 1971) 478–504.

2. G. E. Collins, Subresultants and reduced polynomial remainder sequences, J. ACM, 14 (Jan. 1967) 128–142.

3. N. Jacobson, Basic Algebra I, W. H. Freeman, 1974.

4. D. E. Knuth, Seminumerical Algorithms, The Art of Computer Programming, vol. 2, Addison-Wesley, Reading, Mass., 1969.

5. J. R. Rice, On the conditioning of polynomial and rational forms, Numer. Math, 7 (1965) 426–435.

6. A. D. Rich and D. R. Stoutemeyer, Capabilities of the muMATH-79 Computer Algebra Systems for the Intel-8080 Microprocessor, Symbolic and Algebraic Computation, Lecture Notes in Computer Science no. 72, Springer-Verlag, Berlin, 1979.

7. B. L. van der Waerden, Modern Algebra, vols. 1–2, Ungar, 1953.

8. J. H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice-Hall, Englewood Cliffs, N.J., 1963.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota, 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P. Algebra, Theory of Numbers and their Applications. S.M. Nikol'skii. Proc. of Steklov Inst. of Math., No. 148 (1978). AMS, 1981, vii + 283 pp, $88 (P). [ISBN: 0-8218-3046-5]

General, T(15-16), S*, L*. The Image of Eternity: Roots of Time in the Physical World. David Park. New Amer Library, 1980, x + 149 pp, $5.95 (P). [ISBN: 0-452-00551-5] This eloquent treatise by a philosopher who appreciates Einstein, Dirac, and Feynman won the 1980 Phi Beta Kappa book award in science. It studies the question "What is Time?" and responses to this question from the Greeks to modern physics. Probably suitable as a text in a philosophy seminar. JAS

General, T(13-14: 2). Mathematik für Naturwissenschaftler. Josef Hainzl. Teubner, Stuttgart, 1981, 375 pp, (P). [ISBN: 3-519-22326-0] Differs from earlier edition (TR, October 1974) in having chapters on probability and statistics. JD-B

General, S. Playing Blackjack in Atlantic City. C.R. Chambliss, T.C. Roginski. GBC Pr, 1981, 281 pp, $9.95 (P). [ISBN: 0-89650-830-7] Basic strategy, plus variations based on a card counting system, for winning at blackjack when played according to Atlantic City rules. RSK

Precalculus, T(13: 1). Trigonometry for College Students. Nancy Myers. D van Nostrand, 1981, ix + 418 pp, $15.95 (P). [ISBN: 0-442-31179-6] Standard treatment of trigonometry; includes material on logarithms and polar coordinates. Arranged as units with exercises and self-tests (all answers given at back of book). Simplistic style. JRG

Precalculus, T*(13- 1). Algebra and Trigonometry with Calculators. Marshall D. Hestenes, Richard O. Hill, Jr. Prentice-Hall, 1981, xiii + 555 pp, $19.95. [ISBN: 0-13-021857-X] The standard topics of a precalculus course including algebraic expressions, functions, conic sections, logarithms and exponents, systems of equations, trigonometry, complex numbers, polynomials and sequences. The text requires the use of a scientific calculator and explains its operation. Features lots of worked examples, exercises and word problems. CEC

Precalculus, T(13: 1), L. College Algebra. Robert Ellis, Denny Gulick. HarBrace J, 1981, ix + 514 pp, $17.95. [ISBN: 0-15-507905-0] Essentially a slimmed-down version of College Algebra and Trigonometry by the same authors, this is a straightforward presentation of the traditional topics together with a brief introduction to matrices and linear programming. Most notable (and best) feature is the problem sets which appear to be better than average both in careful selection and in interest level. JS

Education, P. The Mathematical Education of Exceptional Children and Youth: An Interdisciplinary Approach. Ed: Vincent J. Glennon. NCTM, 1981, vii + 408 pp, $28. [ISBN: 0-87353-171-X] Seven essays on teaching mathematics to children with various types of exceptionalities, each co-authored by a mathematics educator and an expert in special education. Treats gifted and talented students and those with special learning and behavior problems as well as children with physical or sensory impairments. Chapters on legal issues, theory of mathematics education and applications to pre-service and in-service programs. MW

Education, P*, L. Teaching Statistics and Probability: 1981 Yearbook. Ed: Albert P. Shulte. NCTM, 1981, x + 246 pp, $13.75. [ISBN: 0-87353-170-1] Collection of 30 articles dealing with the teaching of statistics and probability in grades K-14. Includes a variety of classroom-tested activities. A valuable source book for the up-to-date teacher. RSK

Education, T. An Activity Approach to Elementary Concepts of Mathematics. Douglas B. Smith, William R. Topp. Addison-Wesley, 1981, viii + 293 pp, $9.95 (P). [ISBN: 0-201-07694-2] Workbook to supplement a course in concepts of mathematics for pre-service elementary teachers. Activities familiarize students with ways to use manipulatives on topics through fractions, decimals and informal

geometry. MW

History, L. Emmy Noether, 1882-1935. Auguste Dick. Trans: H.I. Blocher. Birkhäuser Boston, 1981, xiv + 193 pp, $12.95. [ISBN: 3-7643-3019-8] Fascinating account of the great mathematician's life. Includes the obituaries of B.L. van der Waerden, Hermann Weyl, P.S. Alexandrov, and a list of her publications. RJA

Foundations, P, L. Empiricism, Logic, and Mathematics: Philosophical Papers. Hans Hahn. D Reidel Pub, 1980, xix + 139 pp, $12.95 (P); $28.95. [ISBN 90-277-1066-X; 90-277-1065-1] The mathematician Hans Hahn (1880-1934) was one of the main founders and participants in the Vienna Circle. These lucid essays, dating from 1929 to 1934, deal with the philosophy of mathematics in the context of Hahn's larger commitment to the empiricist philosophy of science. GHM

Number Theory, T(17-18), P, L. Multiplicative Number Theory, Second Edition. Harold Davenport. Revised by Hugh L. Montgomery. Grad. Texts in Math., V. 74. Springer-Verlag, 1980, xiii + 177 pp, $19.80. [ISBN: 0-387-90533-2] Basically a reprinting of Davenport's classic (First Edition, TR, October 1970). However, the exposition of the large sieve has been simplified and expanded by Montgomery. A valuable book for any number theorist. SG

Algebra, T*(15: 1), S, L*. A Modern Course on the Theory of Equations. David E. Dobbs, Robert Hanks. Polygonal Pub, 1980, viii + 216 pp, $15. [ISBN: 0-936428-03-1] An introduction to the theory of polynomial equations which uses modern algebra extensively and takes advantage of the availability of inexpensive hand-held calculators. Includes many interesting historical notes and lots of good exercises. The layout of the text could be improved. A welcome addition to the literature. CEC

Algebra, P. Trees. Jean-Pierre Serre. Trans: John Stillwell. Springer-Verlag, 1980, ix + 142 pp, $29.80. [ISBN: 0-387-10103-9] $SL_2(Q_p)$ is an amalgam of two copies of $SL_2(Z_p)$ which in turn is associated with a graph theoretic tree. Any torsion-free subgroup of $SL_2(Q_p)$ acts freely on this tree and hence is isomorphic to the fundamental group of the quotient of this tree by the action of the subgroup; hence the subgroup of SL_2 is free. This book is a translation of the 1977 Arbres, Amalgames, SL_2 which presents a study of the above observation. JAS

Algebra, T(18: 1), S, P, L. Lecture Notes in Mathematics-830: Polynomial Representations of GL_n. J.A. Green. Springer-Verlag, 1980, vi + 118 pp, $9.80 (P). [ISBN: 0-387-10258-2] A modern but fairly accessible discussion of polynomial representations of $Gl_n(K)$, for K an arbitrary infinite field. Based to a large extent on extensions of ideas and techniques developed by Schur in his early work, particularly his 1901 thesis. Bibliography, index. JS

Algebra, P. Lecture Notes in Mathematics-832: Representation Theory II. Ed: V. Dlab, P. Gabriel. Springer-Verlag, 1980, xiv + 673 pp, $40.20 (P). [ISBN: 0-387-10264-7] Proceedings of the conference held at Carleton University, Ottawa, Canada, August 13-25, 1979. (TR, Part I, March 1976.) JAS

Algebra, P. Lecture Notes in Mathematics-844: Groupe de Brauer. Ed: M. Kervaire, M. Ojanguren. Springer-Verlag, 1981, 274 pp, $16.80 (P). [ISBN: 0-387-10562-X] Some of the papers presented in the seminar at Les Plans-sur-Bex, Switzerland, March 16-22, 1980. JAS

Finite Mathematics, T(13: 1). Finite Mathematics. Lawrence E. Spence. Har-Row, 1981, xi + 531 pp, $19.50. [ISBN: 0-06-046369-4] Includes, besides usual topics, brief treatments of graph theory and first order linear difference equations, but no symbolic logic. Ample problem sets and bibliography. Should be readable by students knowing only high school algebra. MB

Calculus, T(13), L. Elementary Applied Calculus. J.S. Ratti. Mariner Pub, 1981, vii + 445 pp, $17.95. [ISBN: 0-936166-05-3] A fairly standard text in applied calculus. A plethora of exercises, but many few "applied" type exercises. PH

Calculus, T(13: 1). Calculus for Business and Economics. Donald R. Byrkit, Shawky E. Shamma. D van Nostrand, 1981, xi + 372 pp, $16.95. [ISBN: 0-442-21305-0] Presents essential material such as supply and demand functions, cost and revenue functions, market equilibrium and break-even analysis. Then presents the calculus needed to deal with business applications. LLK

Real Analysis, T(15-16), S, L. Advanced Calculus: An Introduction to Modern Analysis. William L. Voxman, Roy H. Goetschel, Jr. Pure and Appl. Math., V. 63. Dekker, 1981, xii + 678 pp, $55. [ISBN: 0-8247-6949-X] A voluminous introduction. One third more material than a year sequence would allow. Emphasizes applicability. Sufficient problems. Quite readable. PH

Differential Equations, T(16-17: 1), S, P, L. Advanced Ordinary Differential Equations. Athanassios G. Kartsatos. Mariner Pub, 1980, xii + 185 pp, $24.50. [ISBN: 0-936166-02-9] Studies the main problems of differential equations in the setting of differential systems. Presents and utilizes a number of functional analysis techniques. Introduces and uses real Banach space theory in the context of solutions to differential equations. Many exercises. A well thought out and well written one semester text. PH

Differential Equations, S(18), P. Lecture Notes in Mathematics-841: Nonlinear Evolution Equations--Global Behavior of Solutions. Alain Haraux. Springer-Verlag, 1981, xii + 313 pp, $18 (P). [ISBN: 0-387-10563-8] A rather technical, advanced treatment of various aspects of the Cauchy problem,

emphasizing recent developments in the area of quasi-autonomous dissipative periodic systems. Assumes a substantial background in Banach and Hilbert space theory as well as some familiarity with nonlinear partial differential equations. Bibliography, brief index. JS

Differential Equations, P. Lecture Notes in Mathematics-827: Ordinary and Partial Differential Equations. Ed: W.N. Everitt. Springer-Verlag, 1980, xvi + 271 pp, $16.80 (P). [ISBN: 0-387-10252-3] Proceedings of the fifth conference held in Dundee, Scotland, March 29-31, 1978. JAS

Differential Equations, P. Lecture Notes in Mathematics-846: Ordinary and Partial Differential Equations. Ed: W.N. Everitt, B.D. Sleeman. Springer-Verlag, 1981, xiv + 384 pp, $22 (P). [ISBN: 0-387-10569-7] Proceedings of the sixth conference held in Dundee, Scotland, March 31-April 4, 1980. JAS

Functional Analysis, S(18), P. Sequence Spaces and Series. P.K. Kamthan, Manjul Gupta. Lect. Notes in Pure and Appl. Math., V. 65. Dekker, 1981, xi + 368 pp, $45 (P). [ISBN: 0-8247-1224-2] Advanced treatment of duality theory of pairs of sequence spaces, recent topological developments in sequence space theory, and convergence theory of series in topological vector spaces. PH

Functional Analysis, P. Bases in Banach Spaces II. Ivan Singer. Springer-Verlag, 1981, viii + 880 pp, $78. [ISBN: 0-387-10394-5] This second of three volumes is concerned with generalizations of the idea of a basis in a Banach space. Most of the text is devoted to countable generalizations; many notes and extensive bibliography. (Volume I, TR, December 1971.) SG

Analysis, P, L. Mathematical Theory of Entropy. Nathaniel F.G. Martin, James W. England. Ency. Math. and its Appl., V. 12. Addison-Wesley, 1981, xxi + 257 pp, $29.50. [ISBN: 0-201-13511-6] John von Neumann suggested that Shannon call his mathematical measure of uncertainty 'entropy,' since noone knew what that was, really. This volume attempts to carefully and completely define the concept and show its relationship with several fields of mathematical investigation, notably ergodic theory. Requires abstract measure theory and some real effort to read. TAV

Analysis, P. Lecture Notes in Mathematics-843: Functional Analysis, Holomorphy, and Approximation Theory. Ed: Silvio Machado. Springer-Verlag, 1981, vi + 636 pp, $36 (P). [ISBN: 0-387-10560-3] Proceedings of the seminar held at the Universidade Federal do Rio de Janeiro, Brazil, August 7-11, 1978. JAS

Differential Geometry, T(16-18: 1, 2), L. A First Course in Differential Geometry. Chuan-Chih Hsiung. Wiley, 1981, xvi + 343 pp, $29.95. [ISBN: 0-471-07953-7] An introduction of the geometry of curves and surfaces in three-space using exterior calculus but avoiding tensors. The material, though classical, is presented with attention to relationships between local and global properties in such a way as to allow extensions to more recent results. The book is essentially self contained at the college senior level since it contains the necessary material on topology and differential forms, as well as material on multivariable calculus and linear geometry. JAS

Differential Geometry, T(15-16: 1), S. Differentialgeometrie für Geodäten. Rudolf J. Taschner. Manzsche-Verlag, 1977, 125 pp, (P). [ISBN: 3-214-00000-4] A brief introduction to differential geometry intended chiefly for students of geodesy. Simple exercises. JD-B

Differential Geometry, P. Lecture Notes in Mathematics-836: Differential Geometrical Methods in Mathematical Physics. Ed: P.L. Garcia, A. Pérez-Rendón, J.M. Souriau. Springer-Verlag, 1980, xii + 538 pp, $29.50 (P). [ISBN: 0-387-10275-2] The proceedings of two conferences: the first held at Aix-en-Provence, September 3-7, 1979, the second at Salamanca, September 10-14, 1979. JAS

Differential Geometry, P. Lecture Notes in Mathematics-838: Global Differential Geometry and Global Analysis. Ed: D. Ferus, et al. Springer-Verlag, 1981, xi + 299 pp, $19.50 (P). [ISBN: 0-387-10285-X] Proceedings of the colloquium held at the Technischen Universität Berlin, November 21-24, 1979. JAS

Differential Geometry, T(17-18: 1, 2), L. Differential Geometric Structures. Walter A. Poor. McGraw-Hill, 1981, xiii + 338 pp, $39.95. [ISBN: 0-07-050435-0] This text fills an important gap in the route to geometric understanding. It requires some background in manifold theory (which could easily be added to a course using this text) and then, starting with fibre bundles, it presents connections, Riemannian structures, and so on through sympletic and other more specialized structures. The text is mostly a formal presentation but offers good explanations as well. It presents large quantities of instructive examples and has both a good bibliography and index. However, it is short on problems. JAS

Geometry, T(17: 1, 2), S*, P, L*. Topological Geometry, Second Edition. Ian R. Porteous. Cambridge U Pr, 1981, 486 pp, $22.50 (P); $59.50. [ISBN: 0-521-29839-3; 0-521-23160-4] This second edition is essentially a corrected reprint of the First Edition (TR, January 1971). Although the approach and level of this book is unsuited for undergraduate classes, it contains much that should be better known about the relation between affine geometry and multivariable calculus. JAS

Geometry, S(15-16). Einführung in die Lineare Algebra und Geometrie, Teil 1. Johann Cigler. Manzsche Verlag, 1976, 128 pp, (P). [ISBN: 3-214-00010-1] Elementary real projective geometry developed analytically along lines strongly influenced by Coxeter, Pedoe and Tietz. No exercises or index. JD-B

Topology, P. Points Fixes Pour Les Applications Compactes: Espaces de Lefschetz et la Théorie de l'Indice. Andrzej Granas. Pr U Montreal, 1980, 189 pp, (P). [ISBN: 2-7606-0466-7] Detailed exposition of theory of Lefschetz spaces and (fixed point) indices for non-compact spaces. Includes an introduction to infinite dimensional cohomology with an appendix (in English) by Geba presenting several applications to bifurcation theory. JRG

Topology, T(17-18), L. Classical Topology and Combinatorial Group Theory. John Stillwell. Grad. Texts in Math., V. 72. Springer-Verlag, 1980, xii + 301 pp, $32. [ISBN: 0-387-90516-2] An interesting introduction to topology that focuses on the fundamental group. The subject is presented in an historical context; ideas are motivated well; a fair number of exercises are included. The emphasis on dimensions at most 3 obviates the need for homology. Topics include: graphs and free groups, fundamental groups of complexes, curves on surfaces, knots and braids, 3-dimensional manifolds. SG

Topology, T*(15-17: 1), L. Topics in Topology. Arlo W. Schurle. Elsevier North Holland, 1979, xi + 266 pp, $24.95. [ISBN: 0-444-00285-5] The topics include point set and geometric topology aimed at a general upper level college student. The exposition is pleasantly readable and thoughtful of the student who is not a dedicated mathematician. Many concepts are presented first in a metric context. Numerous exercises, good index, and a useful but not overwhelming bibliography. JAS

Optimization, T(16-17: 1), S, P. Optimality in Nonlinear Programming: A Feasible Directions Approach. A. Ben-Israel, A. Ben-Tal, S. Zlobec. Wiley, 1981, xii + 144 pp, $21.95. [ISBN: 0-471-08057-8] Presents a unified approach to both convex and non-convex programming based on directions of descent. Yields complete characterization of optimality in convex programming and narrows the gap between necessary and sufficient conditions for non-convex programming. Few exercises. JRG

Probability, P. Probability Based on Radon Measures. Tue Tjur. Wiley, 1980, xi + 232 pp, $55.50. [ISBN: 0-471-27824-6] A relatively short and self-contained treatment of the theory of Radon measures as used in probability that avoids appealing to classical "abstract" measure theory. The author provides a compelling argument that a unified Radon measure approach to probability is sufficiently general for probabilistic reasoning. Note the price. TAV

Probability, P*. Martingale Limit Theory and Its Application. P. Hall, C.C. Heyde. Prob. and Math. Stat. Academic Pr, 1980, xii + 308 pp, $36. [ISBN: 0-12-319350-8] The authors develop their thesis that Martingale limit theory deserves a central position in probability theory by using Martingale limit arguments to investigate generalizations of the law of large numbers, the central limit theorem and the law of the iterated logarithm. With these in hand they investigate powerful applications. A very intriguing book. TAV

Probability, T*(16: 1, 2), S, P, L. Introduction to Probability Models, Second Edition. Sheldon M. Ross. Prob. and Math. Stat. Acad Pr, 1980, xi + 376 pp, $21.95. [ISBN: 0-12-598460-X] Revision of the author's 1972 text (TR, March 1973; ER, November 1975). Material on queueing theory has been expanded and placed in a separate chapter. New applications and examples have been added in several other chapters. The style remains the same. RSK

Probability, P. The Geometry of Random Fields. Robert J. Adler. Wiley, 1981, xi + 280 pp, $49.25. [ISBN: 0-471-27844-0] A study of the sample function behavior of random fields and the accompanying geometry. Presumes a graduate level understanding of modern probability. Not surprisingly the theory is much richer and more complex than the one dimensional theory. TAV

Probability, P. Dirichlet Forms and Markov Processes. Masatoshi Fukushima. Math. Lib., V. 23. Elsevier North Holland, 1980, x + 196 pp, $39. [ISBN: 0-444-85421-5] The author attempts to unify the theory of Dirichlet spaces (Beurling and Deny, '59) with that of the Hunt process and the standard process. This further tightens the relationship of classical potential theory and Brownian motion. Based upon his 1975 book (in Japanese) with extensions and new material. TAV

Probability, P. Lecture Notes in Statistics-4: Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. Erik van Doorn. Springer-Verlag, 1981, vi + 118 pp, $9.80 (P). [ISBN: 0-387-90547-2] Technical monograph with main result giving circumstances under which a stochastic process with monotone transition operators is stochastically monotone on intervals of the form (t,∞), with $t > 0$. RSK

Probability, P. Lecture Notes in Mathematics-828: Probability Theory on Vector Spaces II. Ed: A. Weron. Springer-Verlag, 1980, xiii + 324 pp, $19.50 (P). [ISBN: 0-387-10253-1] Proceedings of the conference held at Błażejewko, Poland, September 17-23, 1979. JAS

Statistics, P. Factor Analysis in Chemistry. Edmund R. Malinowski, Darryl G. Howery. Wiley, 1980, ix + 251 pp, $30.50. [ISBN: 0-471-05881-5] Theory, practice and application of factor analysis, "one of the most powerful methods...of chemometrics" (the utilization of mathematical and statistical methods with chemical data). Emphasizes the target-transformation approach as opposed to classical abstract factor analysis. Primary applications are to component analysis, nuclear magnetic resonance, and chromatography. RSK

Statistics, T(13: 1). Statistics, Second Edition. Norma Gilbert. Saunders Coll Pub, 1981, xi + 434 pp, $20.95. [ISBN: 0-03-058091-1] Retains first edition's readability and suitability for self-paced study (TR, March 1977; Extended Review, March 1977). Main changes: (1) sections now correspond well in coverage with 50-minute lectures; (2) each section is now followed by a problem

set, followed by detailed answers to all problems; (3) each chapter ends with exercises, with answers to odd problems. MB

Statistics, P. Spatial Statistics. Brian D. Ripley. Wiley, 1981, x + 252 pp, $29.95. [ISBN: 0-471-08367-4] Methods for analyzing two and three dimensional data. Includes many computer generated diagrams. Mathematically sophisticated. Over 600 references. RJK

Statistics, T(15-16: 1), L. Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data. John A. Cornell. Wiley, 1981, xvii + 305 pp, $30.95. [ISBN: 0-471-07916-2] A self contained treatment, collecting for the first time the techniques used in the design and statistical analysis of mixing models. Contains exercises, solutions and a useful annotated bibliography. TAV

Statistics, T(17-18: 1, 2), P. An Introduction to Multivariate Statistics. M.S. Srivastava, C.G. Khatri. Elsevier North Holland, 1979, xvi + 350 pp, $19.50. [ISBN: 0-444-00302-9] In-depth treatment of multivariate testing. "Basic emphasis is on the maximum likelihood method in testing and on Roy's union-intersection principle in obtaining confidence intervals." Includes up-to-date treatment of the multivariate normal and Wishart distributions, and linear regression models. RSK

Statistics, T(16-17: 1), P. The Analysis of Time Series: An Introduction, Second Edition. C. Chatfield. Chapman & Hall, 1980, xiv + 268 pp, (P). [ISBN: 0-412-22460-7] Minor revision and updating of the author's 1975 text, which had the subtitle Theory and Practice (TR, June-July 1976). Good set of references. RSK

Statistics, P. Optimal Design: An Introduction to the Theory for Parameter Estimation. S.D. Silvey. Chapman & Hall, 1980, viii + 86 pp, $17.95. [ISBN: 0-412-22910-2] Monograph providing most important aspects of the theory of optimal design under the assumption that the model underlying the data is known. Roughly two-thirds devoted to linear theory, particularly linear regression, and one-third to non-linear. RSK

Statistics, T*(16-17: 1, 2), S, P, L*. Applied Regression Analysis, Second Edition. N.R. Draper, H. Smith. Wiley, 1981, xiv + 709 pp, $24.95. [ISBN: 0-471-02995-5] In the Wiley Series in Probability and Mathematical Statistics. Revision and expansion of the authors' well-known 1966 text (TR, January 1967), reflecting advances in the field. Among the new topics are serial correlation of residuals; detection of influential observations; families of transformations; best subset, ridge, principal component, latent root and robust regression; and nonlinear growth models. Good bibliography. RSK

Statistics, T(15-17: 1), S, P. Introduction to Data Analysis and Statistical Inference. Carl N. Morris, John E. Rolph. Prentice-Hall, 1981, xx + 389 pp, $13.95 (P). [ISBN: 0-13-480582-8] Text designed for use in the Rand Graduate Institute of Policy Studies, along with the companion manual STATLIB: A Statistical Computing Library (see below). Topics, chosen to be useful in public policy research, range from elementary descriptive measures to sophisticated multivariate techniques. Organized into four main parts: data analysis, statistical inference for linear models, robustness, and additional topics (Bayes and empirical Bayes inference, sampling methods, and experimental design). RSK

Statistics, S(15-17), P. STATLIB: A Statistical Computing Library. William M. Brelsford, Daniel A. Relles. Prentice-Hall, 1981, xviii + 427 pp, $17.50 (P). [ISBN: 0-13-846220-8] Manual containing documentation for subroutines in STATLIB, a statistical computing system used at Bell Laboratories and the Rand Corporation, consisting of a library of Fortran-callable subroutines and a command system for executing them. (See Introduction to Data Analysis and Statistical Inference above.) RSK

Statistics, S(13-16), P*. Applications, Basics, and Computing of Exploratory Data Analysis. Paul F. Velleman, David C. Hoaglin. Duxbury Pr, 1981, xxi + 354 pp, $10.95 (P). [ISBN: 0-87872-273-4] Presents nine selected techniques of exploratory data analysis, illustrates their applications to real data, and provides Basic and Fortran programs for their implementation (available also in machine-readable form from CONDUIT). Based on Tukey's pioneering 1977 book, Exploratory Data Analysis (TR, August-September 1978). RSK

Statistics, T(15-17: 1), S, P. Time Series and Forecasting: An Applied Approach. Bruce L. Bowerman, Richard T. O'Connell. Duxbury Pr, 1979, xiii + 481 pp, $19.95. [ISBN: 0-87872-218-1] Basics of short-term forecasting of time series. Divided into four parts: forecasting and multiple regression analysis, forecasting time series described by trend and irregular components, forecasting seasonal time series, and the Box-Jenkins methodology. RSK

Computer Programming, S(13-14), L. Practical Basic Programs. Ed: Lon Poole. Osborne/McGraw-Hill, 1980, xi + 171 pp, $15 (P). [ISBN: 0-931988-38-1] Listings of forty very practical programs each with discussion and printout of a sample run. The language is a very general Basic that should run with at most very minor modification on any computer. Programs concern finance, scheduling and networks, statistics, and even musical transposition. JAS

Computer Programming, T?(13-14: 1), S, L. A Primer on Pascal, Second Edition. Richard Conway, David Gries, E.C. Zimmerman. Winthrop Pub, 1981, xii + 430 pp, $12.95. [ISBN: 0-87626-675-8] Designed to teach Pascal as a first language, this mixes discussion of "What is a Pascal program" with a slow-paced introduction to UCSD Pascal and Pascal 6000; contains only references to Jensen and Wirth's Pascal User Manual and Report. The treatment reflects the CDC batch implementation of Pascal 6000

and the PL/I experience of the authors. It makes Pascal seem tedious but elementary, rather than gee-whiz slick. (First Edition, TR, February 1977.) JAS

Computer Programming, T?(13-14: 1), S, L. Structured PL/I Programming with Business Applications. Clarence J. Rockey. Wm C Brown Co, 1981, xv + 522 pp, $13.95 (P). [ISBN: 0-697-08141-9] An elementary text on business programming using a subset of PL/I. It reads like a manual: concepts are not sufficiently developed. Perhaps it was written for students who have been exposed to business programming in another language. Not enough exercises. JL

Computer Programming, S(13-14), L. Some Common Basic Programs, Third Edition. Lon Poole, Mary Borchers. Osborne/McGraw-Hill, 1979, xii + 195 pp, $12.50 (P). [ISBN: 0-931988-06-3] This book describes some 76 common Basic programs, with full listings of codes. The programs are simple enough to be used as examples or exercises for an introductory programming course; and the book can be a valuable teaching aid for such a course. (Second Edition, TR, May 1979.) JL

Computer Programming, T(14-15: 1), L. Induction, Recursion, and Programming. Mitchell Wand. Elsevier North Holland, 1980, xiii + 202 pp, $27.95. [ISBN: 0-444-00322-3] A one-semester course in discrete structures, similar to course B3 of ACM Curriculum 68. A main theme is the use of mathematical induction and loop invariants to prove correctness of (recursive) programs. The semantics of recursive schemes is developed à la Knuth, using first-order structures as a model for data types. Wherever possible lessons are drawn with respect to practical programming tasks. GHM

Computer Programming, P, L. Proceedings International Conference on ALGOL 68. Ed: J.C. van Vliet, H. Wupper. Math. Centre Tracts, No. 134. Math Centrum, 1981, iii + 232 pp, Dfl. 29.40 (P). Papers dealing with various aspects of teaching, implementing and applying the programming language ALGOL 68. Most authors argue for the pedagogical and practical benefits of ALGOL 68 based on its orthogonal design, carefully chosen notation, and clarity. Through this progress report they hope to counteract apprehensions about ALGOL 68 which have prevented its widespread adoption. GHM

Computer Programming, T(15-16), P. The Architecture of Digital Computers. R.G. Garside. Clarendon Pr:Oxford U Pr, 1980, x + 365 pp, $74; $34.50 (P). [ISBN: 0-19-859627-8] Aim is to describe computers and all their common variations at the "architecture" level: below the operating system but above the detailed electronics. Appears to be a quite readable general account without bogging down in the idiosyncracies of any one computer. GHM

Computer Programming, S(13). Karel the Robot. Richard E. Pattis. Wiley, 1981, xiv + 106 pp, $8.95. [ISBN: 0-471-08928-1] An innovative approach for teaching programming. To be used at the beginning of an introductory programming course (recommends 4 to 8 lectures), or as a course in computer appreciation. LLK

Computer Programming, S(13-14), P. H-8 Programming for Beginners. Ron Santore, Don Inman, Bob Albrecht. Dilithium Pr, 1980, x + 194 pp, $8.95. [ISBN: 0-918398-17-7] An introduction to 8080 machine (not assembler) in the framework of the Heathkit H-8 computer. Programming is via the H-8 monitor program which uses direct entry into memory in octal. The presentation is very basic. It starts with octal numbering and architecture and goes patiently through most of the 8080 instructions with short sample programs. Avoids interrupts, restart commands, and certain flags. JAS

Computer Science, T(13: 1). Computing: A Problem-Solving Approach with Fortran 77. T. Ray Nanney. Prentice-Hall, 1981, xiv + 530 pp, $17.95. [ISBN: 0-13-165209-5] A text for a beginning course in computer science. In addition to programming in Fortran it contains chapters on algorithms, programming concepts, communication, and social issues. LLK

Computer Science, T(16-18: 1), S, P, L. Inside Computer Understanding: Five Programs Plus Miniatures. Ed: Roger C. Schank, Christopher K. Riesbeck. Lawrence Erlbaum Assoc, 1981, xii + 386 pp, $29.95. [ISBN: 0-89859-071-X] A very stimulating and readable book on natural language processing using computers, an area of artificial intelligence. A theory of modelling human cognition is presented. Five programs, which are based on the theory and which perform various kinds of language processing through modelling aspects of human cognition, are described in detail. Listings of micro versions of the programs are provided. Accessible to artificial intelligence novices. JL

Computer Science, S(14-17), P, L. The 8086 Book. Russell Rector, George Alexy. Osborne/McGraw-Hill, 1980, xxi + 589 pp, $15 (P). [ISBN: 0-931988-29-2] A manual for the INTEL 8086 microprocessor. It contains detailed descriptions of the CPU architecture and the assembly language; and it provides signal and bus information for those interested in designing a microsystem based on the 8086. JL

Computer Science, T(17-18: 2), S, P. Cryptography, A Primer. Alan G. Konheim. Wiley, 1981, xiv + 432 pp, $34.95. [ISBN: 0-471-08132-9] The modesty of the title is misleading, for this "primer" is more than an introduction. It provides a thorough treatment of the foundations of cryptography and then devotes 100 pages to applications: communications, file security, public key systems and electronic signatures. SS

Systems Theory, P. Lecture Notes in Mathematics-845: Invariance and System Theory: Algebraic and Geometric Aspects. Allen Tannenbaum. Springer-Verlag, 1981, ix + 161 pp, $11.80 (P). [ISBN: 0-387-10565-4] Two-fold purpose of these lectures: (1) To introduce theoretical mathematicians to systems theory and convince them of the interesting and deep problems from pure math to be solved there, and (2) To introduce systems theorists to some of the increasingly relevant ideas from

algebraic geometry, geometric invariant theory and related fields. GHM

Applications (Biology), T(13-14: 1, 2), S, P. Mathematics for the Biological Sciences: From Graphs Through Calculus to Differential Equations. J.C. Newby. Clarendon Pr:Oxford U Pr, 1980, xv + 319 pp, $59; $27 (P). [ISBN: 0-19-859623-5; 0-19-859624-3] Besides the usual content of American texts on calculus for the social and biological sciences this English work contains: trigonometric functions, differences and interpolation, infinite series, numerical solution of differential equations--using Basic--and more. Readable, but not by the dormant. MB

Applications (Biology), S(18), P. Enzyme Mathematics. Jean-Pierre Kernevez. Stud. in Math. and Its Appl., V. 10. North-Holland, 1980, xiii + 262 pp, $41.50. [ISBN: 0-444-86122-X] Within the general context of developing suitable mathematical models for enzymes immobilized within membranes, seven systems are considered, dealing with such topics as enzyme kinetics, glucose pumps, cell assemblages, pattern formations, and wave propagation. Each chapter begins with an expository section followed by some development of the mathematical theory, usually a system of partial differential equations. Extensive bibliography, no index. JS

Applications (Economics), P. Econometric Analysis by Control Methods. Gregory C. Chow. Wiley, 1981, xv + 320 pp, $32.95. [ISBN: 0-471-08706-8] In the Wiley Series in Probability and Mathematical Statistics. Extension of the author's 1975 text, Analysis and Control of Dynamic Economic Systems (TR, November 1975). Divided into four parts: techniques of stochastic control (applied to nonlinear systems of simultaneous equations), economic applications, stochastic control under rational expectations, and stochastic control techniques in continuous time. Assumes background of a first course in econometrics. RSK

Applications (Physics), P. Quantum Fields--Algebras, Processes. Ed: Ludwig Streit. Springer-Verlag, 1980, ix + 444 pp, $40.20. [ISBN: 0-387-81607-0] Proceedings of the "Bielefeld Encounters in Physics and Mathematics II," including several expository papers on stochastic differential calculus and W*-categories. JAS

Applications (Physics), P. Felix Bloch and Twentieth-Century Physics. M. Chodorow, et al. William Marsh Rice U, 1980, xiv + 247 pp, (P). [ISBN: 0-89263-246-1] A volume of physics and biographical essays dedicated to Felix Bloch by some of his former students on the occasion of his seventy-fifth birthday. JAS

Applications (Physics), T(15-16), S. Variational Principles in Dynamics and Quantum Theory. Wolfgang Yourgrau, Stanley Mandelstam. Dover, 1979, xiii + 201 pp, $4 (P). [ISBN: 0-486-63773-5] A Dover republication of the 1968 third edition. The book develops both the history and theory of the principle of least action from the point of variational principles, avoiding Poisson brackets, the d'Alembert principle and other more special topics. JAS

Applications (Physics), T(16-17: 1, 2). Operator Methods in Quantum Mechanics. Martin Schechter. Elsevier North Holland, 1981, xx + 324 pp, $32.50. [ISBN: 0-444-00410-6] A textbook in applied operator theory which uses one-dimensional quantum mechanics to motivate development of the theory. Although some knowledge of Lebesgue integration and complex variables would be useful, main prerequisite is working knowledge of advanced calculus. AO

Applications (Physics), P. Energy Methods in Electromagnetism. P. Hammond. Clarendon Pr:Oxford U Pr, 1981, xiv + 180 pp, $55. [ISBN: 0-19-859328-7] Text develops a method of looking at electromagnetic phenomena from the standpoint of energy (rather than fields) which leads to new computational techniques. AO

Applications (Physics), P*. Operator Algebras and Quantum Statistical Mechanics II: Equilibrium States, Models in Quantum Statistical Mechanics. Ola Bratteli, Derek W. Robinson. Texts and Mono. in Physics. Springer-Verlag, 1981, xi + 505 pp, $46. [ISBN: 0-387-10381-3] An in-depth exposition of applications of C*- and W*-algebraic techniques in the study of macroscopic systems using quantum statistical mechanics. AO

Applications (Physics), P. Lecture Notes in Physics-132: Systems Far from Equilibrium. Ed: L. Garrido. Springer-Verlag, 1980, xv + 402 pp, $27.70 (P). [ISBN: 0-387-10251-5] Proceedings of the conference held at Sitges, Barcelona, Spain in June 1980. JAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Northeast Section Meeting

The summer meeting of the Northeast Section was held June 12-13, 1981 at New England College in Henniker, New Hampshire. Approximately 55 people attended the meeting.

Invited Lectures:

"Mathematics Tomorrow: New Challenges, New Opportunities," by Lynn A. Steen, St. Olaf College.
"Numerical Techniques for Approximating Partial Differential Equations," by Joseph R. Caspar, United Technologies Research Center.
"Some Consequences of the Classification of Finite Simple Groups," by Walter Feit, Yale University.

Short Presentations:

"A Self-Paced Calculus Program that Works," by James E. Ward, Bowdoin College.
"Transparencies for Calculus," by Phil M. Locke, University of Maine.
"A Mathematics Program for Gifted-Talented High School Students," by Jacqueline Mitchell and Robert Jenkins, Portland Public Schools.

At the business meeting a change in the by-laws concerning meeting registration fees was approved.

Pacific Northwest Section

The Pacific Northwest Section held its annual meeting in conjunction with the American Mathematical Society and the Society for Industrial and Applied Mathematics at Lewis and Clark College in Portland, Oregon on Friday and Saturday, June 19 and 20, 1981.

Invited Lectures:

"Why Study Equations over Finite Fields?," by Neal Koblitz, University of Washington.
"The Method of Proof by Example in Number Theory," by Harold M. Stark, University of California at Berkeley. (AMS)
"Non-Linear Hyperbolic Waves," by Andrew Majda, University of California, Berkeley. (AMS)
"The Way It Was," by Ivan Niven, University of Oregon.
"Mathematics Used in Monitoring Mt. St. Helens," by Ansel Johnson, Portland State University.

Short Presentations:

"Mathematics in Pursuit of Fairness," by Roland Lamberson, Humboldt State University.
"Rigid Syllabi and Flexible Playpens," by Richard Montgomery, Southern Oregon State College.
"Use of Video Recordings to Illustrate and Motivate Some Topics in Vector Calculus," by Michael Sequeria, Clark College.
"Logarithmic Scales as a Motivational Gimmick," by Larry Runyon, Bellevue Community College.
"A Portrait of a Mathematician," by Roy Ryden, Humboldt State University.
"Factorization of Sets of Integers," by Calvin T. Long, Washington State University.
"Uniform Distributions a la Niven," by Lawrence E. Eggan, Illinois State University.
"On Constructing the Reals: An Alternative to Cauchy Sequences and Dedekind Cuts," by Eugene Maier, Portland State University.
"Submatrices with Identical Entries," by Charles L. Vanden Eynden, Illinois State Univ. (AMS)
"Number of Positive Solutions of Linear Diophantine Equations," by Burke Zane, California State University at Fresno. (AMS)
"Two Often Considered Divisibility Problems," by Rodney T. Hansen, Montana State University. (AMS)
"Partitions Mod m," by Robert D. Stalley, Oregon State University. (AMS)
"Helping Returning Women Students," by Nancy Cook, University of Washington.
"Strategies for Success with the Turned-Off Math Student," by Dena Patterson, University of California at Berkeley.
"Classroom Stratregies for Dealing with Math Anxiety," by Joyce Duchesneau, Lane Comm. College.
"Math Anxiety, A Joint Counseling-Math Approach," by Sue Kaplan, Western Washington University.

Panel Discussion of Mathematics Contests:

"Putnam Examination," by William Firey, Oregon State University.
"MAA High School Mathematics Contest," by Walter E. Mientka, University of Nebraska.
"Washington Local, Regional & State Contests," by Sue Melchior, Shorecrest High School.
"Oregon Local, Regional & State Contests," by Don Gallagher, Central Oregon Community College.

ERRATUM

Due to printer's error
page C7 and page C8 are transposed.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4*

IS THERE A PERMANENTAL ANALOGUE TO OPPENHEIM'S INEQUALITY?

JOHN CHOLLET
Department of Mathematics, University of British Columbia, Vancouver, B. C. V6T 1W5

1. Preliminaries. We begin by reviewing two concepts from matrix theory: the Hadamard product (also called the Schur product) and permanents. The Hadamard product of two $m \times m$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ is defined as the $m \times m$ matrix $A \circ B$ whose (i, j) entry is $a_{ij}b_{ij}$. The properties of the Hadamard product are discussed at length in **[5]** and more briefly in **[2]**. One of these properties asserts that $A \circ B$ is positive semidefinite hermitian whenever A and B are positive semidefinite hermitian. To define the second concept, we recall that the determinant of the matrix A is given by

$$\det A = \sum_{\sigma \in S_m} \varepsilon(\sigma) \prod_{i=1}^{m} a_{i\sigma(i)}, \tag{1}$$

where σ is an element of the symmetric group S_m and $\varepsilon(\sigma)$ is $+1$ if σ is an even and -1 if σ is an odd permutation. If in (1), we replace all $\varepsilon(\sigma)$'s by $+1$, we obtain the permanent of A; that is,

$$\operatorname{per} A = \sum_{\sigma \in S_m} \prod_{i=1}^{m} a_{i\sigma(i)}.$$

The permanent of a positive semidefinite hermitian matrix is a nonnegative real number. An extensive discussion of permanents can be found in **[3]**.

2. The Problem. In this section, A and B will denote $m \times m$ positive semidefinite hermitian matrices. In **[4]**, Oppenheim obtained the inequality

$$\det(A \circ B) \geqslant \det A \det B, \tag{2}$$

which, by equating B to the identity matrix, becomes Hadamard's inequality:

$$\prod_{i=1}^{m} a_{ii} \geqslant \det A.$$

An analogous inequality for permanents **[1]** is

$$\prod_{i=1}^{m} a_{ii} \leqslant \operatorname{per} A. \tag{3}$$

Is there a permanental analogue to Oppenheim's inequality? If the determinants appearing in (2) are replaced by permanents and the sense of the inequality is reversed, one obtains

$$¿ \qquad \operatorname{per}(A \circ B) \leqslant \operatorname{per} A \operatorname{per} B. \qquad ? \tag{4}$$

Note that (4) implies (3). The problem is to decide whether inequality (4) is true.

3. An Equivalent Inequality. Let $\bar{A} = [\bar{a}_{ij}]$, where $\bar{a}_{ij}$ is the complex conjugate of a_{ij}. If A is positive semidefinite, so is $\bar{A}$; and $\operatorname{per} A = \operatorname{per} \bar{A}$. By setting B equal to $\bar{A}$, inequality (4) becomes

$$¿ \quad \operatorname{per}(A \circ \bar{A}) \leqslant (\operatorname{per} A)^2. \quad ? \tag{5}$$

Now assume the truth of (5) for all positive semidefinite hermitian matrices A. Note that

$$\operatorname{per}(A \circ \bar{A}) = \sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}|^2.$$

Then,

$$0 \leqslant \operatorname{per}(A \circ B) = \sum_{\sigma \in S_m} \prod_{i=1}^{m} a_{i\sigma(i)} b_{i\sigma(i)} \leqslant \sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}| |b_{i\sigma(i)}|$$

$$\leqslant \left(\sum_{\sigma \in S_m} \prod_{i=1}^{m} |a_{i\sigma(i)}|^2 \right)^{1/2} \left(\sum_{\sigma \in S_m} \prod_{i=1}^{m} |b_{i\sigma(i)}|^2 \right)^{1/2} = \left(\operatorname{per}(A \circ \bar{A})\right)^{1/2} \left(\operatorname{per}(B \circ \bar{B})\right)^{1/2},$$

$$\operatorname{per}(A \circ B) \leqslant \operatorname{per} A \operatorname{per} B.$$

To get the second line, we applied the Cauchy-Schwarz inequality. Therefore, inequalities (4) and (5) are equivalent.

4. Example. If

$$A = \begin{bmatrix} 2 & -i & 1 \\ i & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 3 & 2i \\ 3 & 6 & i \\ -2i & -i & 5 \end{bmatrix},$$

then

$$A \circ B = \begin{bmatrix} 8 & -3i & 2i \\ 3i & 12 & 0 \\ -2i & 0 & 15 \end{bmatrix}.$$

It is clear that all 1×1, 2×2 and 3×3 principal subdeterminants of the hermitian matrices A, B and $A \circ B$ are nonnegative. Therefore, by a known theorem, these matrices are positive semidefinite hermitian. A short calculation now shows that $\operatorname{per} A = 17$, $\operatorname{per} B = 205$, $\operatorname{per}(A \circ B) = 1623$ and $\operatorname{per} A \operatorname{per} B = 3485$.

References

1. M. Marcus, The Hadamard theorem for permanents, Proc. Amer. Math. Soc., 15 (1964) 967–973.
2. M. Marcus and H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber and Schmidt, Boston, 1964.
3. H. Minc, Permanents, Addison-Wesley, Reading, Mass., 1978.
4. A. Oppenheim, Inequalities connected with definite hermitian forms, J. London Math. Soc., 5 (1930) 114–119.
5. G. P. H. Styan, Hadamard products and multivariate statistical analysis, Linear Algebra and Appl., 6 (1973) 217–240.

ANSWERS TO "PHOTOS" ON PAGE 33

Top left: E. Artin; top right: S. Banach; bottom left: A. Grothendieck; bottom right: C. Shannon. Artin, Banach, and Shannon would probably be recognized by their friends, but Grothendieck probably not. Grothendieck's picture was taken in 1954 ± 1.

NOTES

EDITED BY J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

A SIMPLE PROOF OF THE FOUR-SQUARES THEOREM

CHARLES SMALL
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

The fact that every positive integer is a sum of four squares was known long before it was first proved (in 1770, by Lagrange) and it has been given many different proofs since. The purpose of this note is to give a short proof that makes a lovely and convincing application in undergraduate algebra courses.

In this approach the four-squares theorem follows from a certain factorization theorem for 2×2 matrices over the ring $Z[i]$ of Gaussian integers. The proof is due to M. Newman (see [**1**, p. 813]). The version given here is a simplification of that given by G. M. Bergman [**2**].

The starting point is a well-known result:

LEMMA. *Let p be a prime and let Z_p denote the ring of integers* $\bmod p$. *Then every element $r \in Z_p$ is a sum of two squares.*

Proof. We may assume $p \neq 2$. Put $S_1 = \{x^2 \mid x \in Z_p\}$ and $S_2 = \{r - y^2 \mid y \in Z_p\}$. Then S_1 and S_2 each has $(p+1)/2$ elements. Therefore $S_1 \cap S_2$ is nonempty; so r is a sum of two squares.

REMARK. The same proof shows more generally that in any finite field F any equation $ax^2 + cy^2 = r$ $(a, c \neq 0)$ has solutions: let $S_1 = \{ax^2 \mid x \in F\}$ and $S_2 = \{r - cy^2 \mid y \in F\}$. The reader is invited to prove, more generally still, that $ax^2 + bxy + cy^2 = r$ has solutions if $b^2 - 4ac \neq 0$. Thus "every nondegenerate binary quadratic form over a finite field is universal."

COROLLARY. *Let n be squarefree and let Z_n denote the ring of integers* $\bmod n$. *Then every element $r \in Z_n$ is a sum of two squares.*

Proof. This follows from the lemma with the aid of the Chinese Remainder Theorem: n, being squarefree, is a product of distinct primes, say $n = p_1 \cdots p_k$. Then $Z_n \cong Z_{p_1} \times \cdots \times Z_{p_k}$, and therefore writing r as a sum of two squares in each Z_{p_i} writes r as a sum of two squares in Z_n.

REMARK. The converse of the corollary is false: every element of Z_{25} is a sum of two squares. The problem this suggests—given n, determine the smallest m such that every element of Z_n is a sum of m squares—is solved in [**3**, §3], and the generalization from squares to kth powers is handled in [**3**] and [**4**].

Now, to see that a positive integer n is a sum of four squares, we may assume that n is squarefree; for if $n = a^2 n'$ (n' squarefree) and if $n' = w^2 + x^2 + y^2 + z^2$ then $n = (aw)^2 + (ax)^2 + (ay)^2 + (az)^2$. Once n is squarefree, we use the corollary above (with $r = -1$) to find integers c, d, and m with $-1 = c^2 + d^2 - mn$. Now consider the matrix

$$A = \begin{pmatrix} n & c + di \\ c - di & m \end{pmatrix}$$

(where of course $i = \sqrt{-1}$). Observe that $\det A = nm - c^2 - d^2 = 1$. The factorization theorem for such matrices is as follows:

THEOREM. *Let*

$$A = \begin{pmatrix} n & c + di \\ c - di & m \end{pmatrix}$$

where c, d, m, n are integers and $n > 0$, and assume $\det A = nm - c^2 - d^2 = 1$. *Then* $A = BB^*$ *for some* 2×2 *matrix B over $Z[i]$, where B^* denotes the conjugate transpose of B.*

The four squares theorem is an immediate consequence:

$$\begin{pmatrix} n & * \\ * & * \end{pmatrix} = \begin{pmatrix} w + xi & y + zi \\ * & * \end{pmatrix} \begin{pmatrix} w - xi & * \\ y - zi & * \end{pmatrix}$$

gives $n = w^2 + x^2 + y^2 + z^2$.

To prove the factorization theorem we do induction on $c^2 + d^2$. If $c^2 + d^2 = 0$ the hypotheses force $A = I$, in which case $B = I$ will do. For the rest of the proof we assume, therefore, that $c^2 + d^2 > 0$; i.e., c and d are not both 0.

Since m is also positive ($nm = 1 + c^2 + d^2$), there are two cases to consider: $0 < n \leqslant m$ or $0 < m \leqslant n$.

Case 1. $0 < n \leqslant m$. Let $A' = MAM^*$ where

$$M = \begin{pmatrix} 1 & 0 \\ x - yi & 1 \end{pmatrix}$$

and x and y are integers to be specified momentarily. Then

$$A' = \begin{pmatrix} n & c' + d'i \\ c' - d'i & * \end{pmatrix}$$

with $c' = c + nx$, $d' = d + ny$, and $\det A' = 1$ because $\det M = \det M^* = 1$. If we can choose x and y to ensure $c'^2 + d'^2 < c^2 + d^2$, we are done: for then $A' = CC^*$ by induction, and putting $B = M^{-1}C$ gives $A = M^{-1}A'(M^*)^{-1} = M^{-1}(CC^*)(M^{-1})^* = (M^{-1}C)(M^{-1}C)^* = BB^*$.

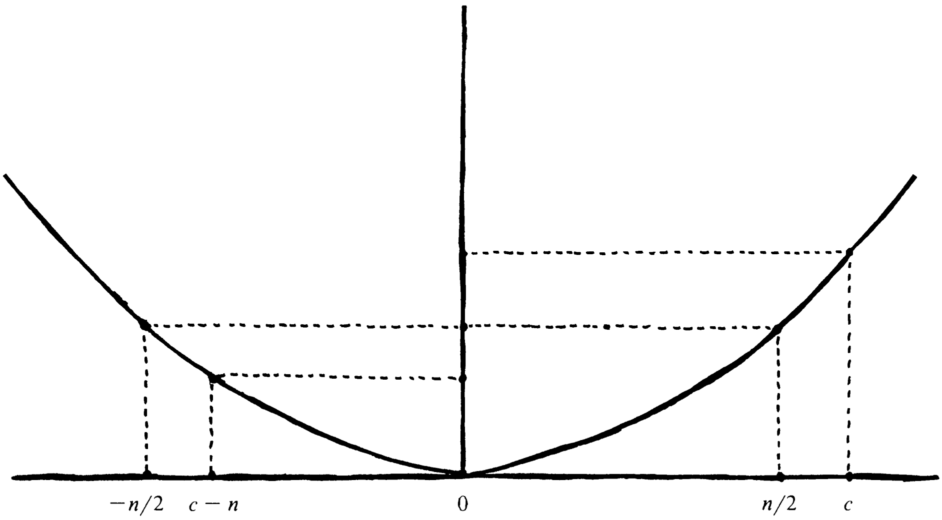


FIG. 1

Now as long as $c > n/2$, we can simply choose $x = -1, y = 0$ (see Fig. 1); then $c'^2 = (c - n)^2 < c^2$ and $d'^2 = d^2$, so that $c'^2 + d'^2 < c^2 + d^2$. Similarly, if $c < -n/2$ we can take $x, y = 1, 0$; if $d > n/2$ take $x, y = 0, -1$; and if $d < -n/2$ take $x, y = 0, 1$.

Hence, to complete the proof (in Case) we need only show that we are necessarily in one of these four situations; that is, $|c| > n/2$ or $|d| > n/2$. If $n = 1$ this is clear since c and d are not both 0. If $n > 1$, suppose the contrary: $|c| \leqslant n/2$ and $|d| \leqslant n/2$. Then, since $0 < n \leqslant m$, we have $n^2 \leqslant nm = c^2 + d^2 + 1 \leqslant (n/2)^2 + (n/2)^2 + 1 = n^2/2 + 1 < n^2$, a contradiction. This completes the proof in Case 1.

For Case 2 ($0 < m \leqslant n$) the argument is similar. Let

$$A' = MAM^* \text{ where } M = \begin{pmatrix} 1 & x+yi \\ 0 & 1 \end{pmatrix};$$

then

$$A' = \begin{pmatrix} * & c'+d'i \\ c'-d'i & m \end{pmatrix} \quad \text{with} \quad c' = c + mx, \quad d' = d + my.$$

Again it suffices to find x and y so that $c'^2 + d'^2 < c^2 + d^2$, and again we can take $x, y = \pm 1, 0$ or $0, \pm 1$, since $|c| > m/2$ or $|d| > m/2$. Done!

Query. What kind of representation theorems for integers can be obtained by studying matrices over $Z[\omega]$ (ω a primitive cube root of 1) instead of $Z[i]$? Newman's book [5] has some remarks bearing on this question; see Chapter 11, especially page 215.

References

1. O. Taussky, Sums of squares, this MONTHLY, 77 (1970) 805–830.
2. Response to Query #26, Notices of A.M.S., 21 (1974) 220–221.
3. C. Small, Waring's Problem mod n, this MONTHLY, 84 (1977) 12–25.
4. ______, Solution of Waring's Problem mod n, this MONTHLY, 84 (1977) 356–359.
5. M. Newman, Integral Matrices, Academic Press, New York-London, 1972.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 44859.

TEACHING AIDS

MOSS E. SWEEDLER
Department of Mathematics, Cornell University, Ithaca, NY 14853

Here are two devices that have proved useful in teaching first-year calculus. The first helps students remember the popular values of the trigonometric functions. The key is the pattern

$$\frac{\sqrt{0}}{2} \quad \frac{\sqrt{1}}{2} \quad \frac{\sqrt{2}}{2} \quad \frac{\sqrt{3}}{2} \quad \frac{\sqrt{4}}{2}.$$

The student still has to remember that these are the values of sine at 0°, 30°, 45°, 60°, 90° and how to get cosine and the other values of trigonometric functions.

The second is how to convince a class that .999... equals 1 and is not the number just before 1. Write

$$\frac{1}{3} = .333\ldots$$

on the board, and everyone will agree. Then tell the class to multiply both sides by three. This usually convinces 99.999... percent of the students.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Diophantine Equations

S 31 [1980, 403]. *Proposed by Leo J. Alex, SUNY College at Oneonta.*

In each of the following, find all solutions in nonnegative integers a, b, c:
(i) $1 + 5^a = 2 \cdot 3^b + 3 \cdot 2^c$;
(ii) $1 + 2^a = 4 \cdot 3^b + 5^c$.

Solution by Lorraine L. Foster, California State University, Northridge. (i) The only solutions are $(a, b, c) = (2,0,3)$ or $(5,3,10)$. For let (a, b, c) be another solution. Suppose $b = 0$; clearly $c \geqslant 4, 5^a \equiv 1 \pmod{16}, 4 \mid a, 3 \cdot 2^c \equiv 0 \pmod{13}$, a contradiction. Hence $b \geqslant 1$. Working mod 5, one sees that $b \neq 1$ and that $(b, c) \equiv (0,3), (2,0)$, or $(3,2)$ mod 4. Thus $b > 1$.

Using mod 3, a is odd. Hence, using mod 8, either $c = 2$ and b is even or $c \geqslant 3$ and b is odd. Thus, using mod 63, $(a, b, c) \equiv (1,5,3), (5,1,2)$ $(b > 1)$ or $(5,3,4) \pmod 6$. Hence using mod 13, $(a, b, c) \equiv (5,3,10) \pmod{12}$. If $b = 3$ then $a > 5, 3 \cdot 2^c \equiv -53 \pmod{15625}$, $c \equiv 10010 \pmod{12500}$, $c \equiv 22510 \pmod{37500}$, $5^a \equiv 53 + 3 \cdot 2^{22510} \equiv 33290 \pmod{37501}$, a contradiction. Thus $b > 3$. Using mod 27, 19 successively, $(a, b, c) \equiv (5,3,10) \pmod{18}$. Finally, using mod 81 we conclude that $a \equiv 23 \pmod{54}$, which yields a contradiction mod 109.

(ii) The only solutions are $(a, b, c) = (2,0,0), (3,0,1), (4,1,1), (7,0,3)$, or $(12,5,5)$. For, let (a, b, c) be another solution. Suppose $b = 0$. Clearly $a \geqslant 10$, $5^c \equiv -3 \pmod{1024}$, $c \equiv 163 \pmod{256}$. Thus $2^a \equiv 246 \pmod{257}$, a contradiction. Hence $b \geqslant 1$. If $b = 1$ then $c \geqslant 2$, $2^a \equiv 11 \pmod{25}$, $a \equiv 16 \pmod{20}$, $0 \equiv 9 + 5^c \pmod{31}$, a contradiction. Thus $b > 1$, $a \geqslant 3$, so that, using mod 8, c is odd. Using mod 9, 7 successively, we have $(a, b, c) \equiv (0,5,5)$ or $(4,0,3) \pmod 6$. Hence, using mod 5, 13 successively, $(a, b, c) \equiv (0,5,5) \pmod{12}$. Suppose $b = 5$. Clearly $c \geqslant 6, 2^a \equiv 96 \pmod{125}$, $a \equiv 12 \pmod{100}$. Thus, using mod 101, $c \equiv 5 \pmod{25}$. Also, $2^a \equiv 971 \pmod{15625}$, $a \equiv 7512 \pmod{12500}$, $5^c \equiv 18379 \pmod{37501}$, $c \equiv 4475 \pmod{18750}$, a contradiction. Hence $b > 5$. Using mod 27, 19 successively, we have $(a, b, c) \equiv (12,5,5) \pmod{36}$. Also, using mod 81, 109, we have $(a, b, c) \equiv (12,5,5) \pmod{108}$. Further using mod 243, 163, $(a, b, c) \equiv (12,5,5) \pmod{324}$. Finally, using mod 729, 487 successively, we have $(a, b, c) \equiv (12,5,167) \pmod{486}$, which is impossible mod 1459.

Also solved by Albert S. Rosenthal and the proposer. Part (i) was solved by L. Kuipers (Switzerland).

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by May 31, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2920. *Proposed by J. O. Shallit, University of California, Berkeley.*

Let $f(x) = x^n - x^{n-1} - x^{n-2} - \cdots - x - 1$, $n \geqslant 2$.

(a) Show that the discriminant of f is given by

$$\operatorname{disc}(f) = (-1)^{n(n+1)/2}\left[\frac{(n+1)^{n+1} - 2^{n+1}n^n}{(n-1)^2}\right].$$

(b) Since f has integer coefficients, $\operatorname{disc}(f)$ must be an integer; hence

$$\frac{(n+1)^{n+1} - 2^{n+1}n^n}{(n-1)^2}$$

must be an integer. Show this in another way.

E 2921. *Proposed by P. R. Halmos, Indiana University.*

Evaluate

$$\limsup_{n\to\infty} |(2+3i)^n - (2-3i)^n|^{1/n}.$$

The problem looks frightening in this form. To make it look less so, put

$$r(a,b) = \limsup_{n\to\infty} |a^n - b^n|^{1/n}$$

for each pair of complex numbers a, b. If $a = b$, then $r(a,b) = 0$; what is it if $a \neq b$?

E 2922. *Proposed by Roger Cuculière, Paris, France.*

Let $\{F(n)\}$ be the Fibonacci sequence: $F(0) = 0$, $F(1) = 1$, $F(n) = F(n-1) + F(n-2)$. Find

$$\sum_{m=0}^{\infty} \frac{1}{F(2^m)}.$$

E 2923. *Proposed by P. Erdős, Hungarian Academy of Sciences, and Claudia Spiro, University of Illinois.*

Let $1 < a_1 < a_2 < \cdots$ be an infinite sequence of integers. Prove that

$$\sum_{n=1}^{\infty} \frac{2^{a_n}}{a_n!}$$

is irrational.

E 2924*. *Proposed by Jack Garfunkel, Flushing, N.Y.*

Triangle $A_1A_2A_3$ is inscribed in a circle; the medians through $A_1\,[A_2]$ meet the circle again at $M_1\,[M_2]$. The angle bisectors through $A_1\,[A_2]$ meet the circle again at $T_1\,[T_2]$. Prove or disprove: $|A_1M_1 - A_2M_2| \leqslant |A_1T_1 - A_2T_2|$.

E 2925. *Proposed by T. Keller, Honolulu, Hawaii.*

For how many integers in base b ($b \geqslant 2$) is the integer equal to the sum of the squares of its digits?

SOLUTIONS OF ELEMENTARY PROBLEMS

A Completely Monotonic Function

E 2845 [1980, 577]. *Proposed by Douglas Hensley, Texas A & M University.*

Let $f(x) = (x^2 - 1)^{1/2}$, $x > 1$. Prove that $f^{(n)}(x) > 0$ for odd n and $f^{(n)}(x) < 0$ for even $n > 0$.

Solution by O. P. Lossers Jr., Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands. The graph of f being the northeastern part of the hyperbola $x^2 - y^2 = 1$, we see that $f(x) > 0, f'(x) > 0, f''(x) < 0$, $(x > 1)$ (the last inequality follows from convexity). We proceed by induction. Assume $n \geqslant 3$.

$$f^2(x) \equiv x^2 - 1 \qquad (x > 1)$$

Differentiation n times in succession yields

$$2ff^{(n)} + \sum_{k=1}^{n-1} \binom{n}{k} f^{(k)} f^{(n-k)} \equiv 0.$$

$f(x)$ is positive. All terms $f^{(k)}(x)f^{(n-k)}(x)$ have the same sign by induction and, hence, the sign of $f'(x)f^{(n-1)}(x)$. Also, $f'(x)$ is positive. Therefore

$$\operatorname{Sgn} f^{(n)}(x) = -\operatorname{Sgn} f^{(n-1)}(x). \qquad \square$$

A function f is completely monotonic (cm) (on an interval of the real line) if $(-1)^n d^n f/dx^n > 0$. Any cm function on $x > a$ can be represented as a Laplace transform $\int_0^\infty e^{-(x-a)s} d\alpha(s)$, with α nondecreasing.

See D. V. Widder, *The Laplace transform*, pp. 160–162. M. F. Kruelle noted that f is completely monotonic if $f > 0, f' > 0, f'' < 0$, $[ff']'' = 0$. M. E. Muldoon referred to S. Bochner, Duke Math. J., 3 (1937) 488–502. I. J. Schoenberg noted that $g = x - \sqrt{x^2 - 1}$ is cm on $x > 1$. O. G. Ruehr found $\beta(s)$ in the representation $g = \int_0^\infty e^{-(x-1)s} d\beta(s)$.

Also solved by 69 other solvers, including the proposer.

Permuted Gram Matrix

E 2849 [1981, 671]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let $f_1, \ldots, f_n$ be n real linearly independent square integrable functions defined on an interval I. Let $G = (g_{ij})$ denote their Gram matrix, i.e., $g_{ij} = \int f_i(x) f_j(x)\,dx$. Let Q denote the matrix obtained by interchanging simultaneously m pairs of rows of G ($2m \leqslant n$). Show that the eigenvalues of Q are all real.

Solution by Ron Adin, Haifa, Israel; H. Kestelman, University College, London; M. F. Kruelle, student, Air Force Institute of Technology; Daniel B. Shapiro, Ohio State University; and Henry Wolkowicz, University of Alberta. We note that $Q = PG$, where P is a permutation matrix, $P^2 = \mathrm{Id}$. Since G is positive definite, G has a positive definite hermitian square root, H ($H^2 = HH^* = G$). Thus $HQH^{-1} = HPGH^{-1} = HPH = HPH^*$. But $P = P^*$, and the result follows.

Also solved by the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by May 31, 1982. The solver's full post-office address should be on each sheet.

6374*. *Proposed by Lee Whitt, Texas A & M University.*

Suppose $f(x)$, $-\infty < x < \infty$, is a real valued function such that both $(f(x))^2$ and $(f(x))^3$ are C^∞. Must f be C^∞?

6375. *Proposed by Chang Gengzhe, University of Utah.*

Set

$$S_n = 1 + \frac{n-1}{n+2} + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right) + \cdots + \left(\frac{n-1}{n+2}\right)\left(\frac{n-2}{n+3}\right)\cdots\left(\frac{1}{2n}\right).$$

Show that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} S_n = \frac{\sqrt{\pi}}{2}.$$

6376. *Proposed by I. J. Schoenberg, University of Wisconsin.*

Let a_n $(n = 1, 2, \ldots)$ be reals. Show that

$$\lim_{n\to\infty} a_n \sum_1^n a_i^2 = 1 \quad \text{implies that} \quad \lim_{n\to\infty} (3n)^{1/3} a_n = 1.$$

SOLUTIONS OF ADVANCED PROBLEMS

Collineations of Projective Spaces

6236 [1978, 770; 1981, 69]. *Proposed by Antal E. Fekete, Memorial University of Newfoundland.*

We say that two endomorphisms of the complex vector space C^n are of the same type if there is a bijection between their respective sets of eigenvalues which maps the Jordan normal form of one endomorphism into that of the other. Find a formula determining the number of different endomorphism types of C^n. Define what is meant by an endomorphism type of the real vector space R^n and determine their number.

Solution by the proposer. Let $p(n)$ denote the number of partitions of the positive integer n. $1i_1 + 2i_2 + \cdots + ni_n = n$ is such a partition having i_k summands equal to k; if we replace each summand k by a partition of k, for $k = 1, 2, \ldots, n$, then we obtain a *double partition* of n. It is clear that the number $c(n)$ of double partitions of n is given by the formula

$$c(n) = \sum_{1i_1 + \cdots + ni_n = n} \binom{p(1) + i_1 - 1}{i_1}\binom{p(2) + i_2 - 1}{i_2}\cdots\binom{p(n) + i_n - 1}{i_n}$$

where the summation is extended to all partitions of n; e.g., $c(1) = 1$, $c(2) = 3$, $c(3) = 6$, $c(4) = 14$, $c(5) = 27$, $c(6) = 58$, $c(7) = 111$, $c(8) = 223$, $c(9) = 424$, $c(10) = 817$.

The number $c(n)$ of double partitions of n also furnishes the number of endomorphism types of $\mathbb{C}^n$. Indeed, every endomorphism has exactly n eigenvalues which, taken with multiplicities, determine a partition of n. The endomorphism also has exactly n eigenvectors (if the elements of a maximal set of linearly independent eigenvectors are counted with their multiplicities). Since each eigenvec-

tor belongs to a well-defined eigenvalue, the endomorphism also determines a double partition of n. Two endomorphisms will determine the same double partition if, and only if, they have the same type, and every double partition of n gives rise to an endomorphism type, making the correspondence between the set of endomorphism types of $\mathbb{C}^n$ and the set of double partitions of n bijective.

Two endomorphisms of the vector space $\mathbb{R}^n$ are said to have the same type if there is a bijection between their respective sets of eigenvalues and eigenvectors preserving all multiplicities and making complex conjugate pairs correspond. In order to find the number $r(n)$ of endomorphism types of $\mathbb{R}^n$, we have to separate the complex conjugate pairs of eigenvectors from the real eigenvectors by splitting off an even dimensional subspace of $\mathbb{R}^n$ in every possible way. Thus we see that

$$r(2k+1) = c(2k+1) + c(1)c(2k-1) + c(2)c(2k-3) + \cdots + c(k)c(1),$$
$$r(2k) = c(2k) + c(1)c(2k-2) + c(2)c(2k-4) + \cdots + c(k);$$

e.g., $r(1) = 1$, $r(2) = 4$, $r(3) = 7$, $r(4) = 20$, $r(5) = 36$, $r(6) = 87$, $r(7) = 162$, $r(8) = 355$, $r(9) = 666$, $r(10) = 1367$.

Our definition of endomorphism types does not distinguish between zero and nonzero eigenvalues. This leads to the anomalous result that two endomorphisms of the same type may have different ranks; in fact, one may be an automorphism while the other is not. We can remedy this by sharpening our definitions. Two endomorphisms are said to have the same *strong type* if there is a bijection between their respective sets of eigenvalues and eigenvectors which preserves all multiplicities and the zero eigenvalue (and, in the case of $\mathbb{R}^n$, it also preserves complex conjugate pairs). Denoting the number of strong endomorphism types of $\mathbb{C}^n$ by the symbol $C(n)$, we have that

$$C(n) = c(n) + c(n-1)p(1) + c(n-2)p(2) + \cdots + c(1)p(n-1) + p(n),$$

e.g., $C(1) = 2$, $C(2) = 6$, $C(3) = 14$, $C(4) = 34$, $C(5) = 74$, $C(7) = 342$, $C(8) = 713$, $C(9) = 1439$, $C(10) = 2881$.

In order to find the number $R(n)$ of strong endomorphism types of $\mathbb{R}^n$, we notice that the complex conjugate pairs of eigenvectors cannot belong to the zero eigenvalue. Therefore,

$$R(2k+1) = r(2k+1) + c(1)r(2k-1) + c(2)r(2k-3) + \cdots + c(k)r(1),$$
$$R(2k) = r(2k) + c(1)r(2k-2) + c(2)r(2k-4) + \cdots + c(k);$$

e.g., $R(1) = 2$, $R(2) = 7$, $R(3) = 16$, $R(4) = 43$, $R(5) = 94$, $R(6) = 222$, $R(7) = 470$, $R(8) = 1029$, $R(9) = 2115$, $R(10) = 4401$.

We note in passing that, *mutatis mutandis*, the same formulas will also furnish the number of automorphism types of $\mathbb{C}^n$ and $\mathbb{R}^n$, where the strong automorphism types treat the eigenvalue 1 as distinguished.

Poisson Random Variables

6295 [1980, 309]. *Proposed by N. Bromberg, Rutgers University.*

Let $\{X_j\}_{j=0}^{\infty}$ be a sequence of independent, identically distributed Poisson random variables with mean $\lambda < 1$. Let $A_k = \{X_0 + \cdots + X_k \leqslant k\}$. Show that $P\{\cap_{k=0}^{\infty} A_k\} = 1 - \lambda$.

Solution by Howard Weiner, University of California at Davis. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of identically distributed nonnegative integer valued random variables with mean $\lambda < 1$. Let $S_n = \sum_{l=1}^{n} X_l$ and $A_n = \{S_n < n\}$. We show $P\{\cap_{n=1}^{\infty} A_n\} = 1 - \lambda$.

For $r < n$, $E[X_r \mid S_n] = S_n/n$ and hence $E[S_r/r \mid S_n] = S_n/n$. That is, $\{S_n/n\}_{n=1}^{\infty}$ is a backward martingale. On the set $\{S_n < n\}$, let $T_n = \{\text{largest } k,\ 1 \leqslant k \leqslant n, \text{ such that } S_k/k \geqslant 1\}$ and set

$T_n = 0$ otherwise. Since $\{X_n\}$ are nonnegative integer valued, it follows that $S_{T_n}/T_n = 1$ for $T_n > 0$. Let $T = \{\text{largest } k \geq 1 \text{ such that } S_k \geq k\}$ and set $T = 0$ otherwise. The strong law of large numbers implies that $T < \infty$ a.s. Then on $\{S_n < n\}$,

$$P[T_n > 0 \mid S_n] = E\left\{\frac{S_{T_n}}{T_n}; T_n > 0 \mid S_n\right\} = \frac{S_n}{n}.$$

Let $n \to \infty$ to obtain by the strong law of large numbers

$$P[T > 0] = \lim_{n\to\infty} P[T_n > 0 \mid S_n] = \lim_{n\to\infty} \frac{S_n}{n} = \lambda, \text{ a.s.}$$

Hence

$$P\left[\bigcap_{n=0}^{\infty} A_n\right] = P[T = 0] = 1 - \lambda.$$

The basic argument appears in, e.g., Y. S. Chow and H. Teicher, *Probability Theory*, Springer-Verlag, New York, 1978, p. 238.

Also solved by K. L. Bernstein, N. J. Boyneton & O. G. Ruehr, L. E. Clarke (England), S. W. Dharmathikari & T. B. Paine, E. Hertz, P. Humblet & M. Pinsky, J. H. B. Kemperman, I. I. Kotlarski & J. Robertson, L. Sennott, M. Skalsky, J. C. Smit (Netherlands), L. Takács, and the proposer.

REVIEWS

EDITED BY JOHN H. EWING

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE.

A Problem Solving Approach to Mathematics for Elementary School Teachers. By Rick Billstein, Shlomo Libeskind, and Johnny W. Lott. Benjamin/Cummings, Menlo Park, Calif., 1981. xvi + 568 + appendices and index.

DONALD R. KERR, JR.

Department of Mathematics, Indiana University, Bloomington, IN 47405

The mathematics training of elementary teachers is one of the most serious problems facing mathematicians today. Granted there are many other groups of students who are not learning mathematics, but just think of the multiplier effect: if an elementary teacher has only a five-year career, he or she (usually she) will teach a full year of mathematics to about 150 students. Of course, many teachers have much longer careers.

Fortunately, the mathematics goals for elementary school teachers do not have to be too ambitious. Virtually every elementary school teacher uses a mathematics textbook; all we must do is prepare them to use the textbook effectively. Certainly the teacher needs competence in arithmetic, and for most of today's texts she needs just a smattering of knowledge from elementary geometry, probability, and descriptive statistics. Equally important, the teacher should have comfort and confidence in dealing with mathematics. (Otherwise, discomfort will be communicated to children, only the most limited mathematical goals will be sought, and any curriculum innovations will be threatening and will be resisted.) In addition to limited competence and confidence, a certain amount of perspective is important. Teachers do make curricular

decisions which often take the form of omitting or emphasizing certain material. If a teacher lacks perspective on how the material in the curriculum fits together, she will likely make poor decisions.

If the goals are limited, as I claim, then where is the problem? To start with, competence and confidence can be antithetical goals. Promoting competence usually requires enforcing standards, which involves testing and realistic grading. Confidence usually grows out of experiences of success. Moreover, as a group, prospective elementary teachers are not a particularly strong group of mathematics students. There are some strong students among them, but there are also many students with modest mathematics backgrounds. The teacher of a mathematics class for elementary teachers often walks a narrow line between enforcing standards and building confidence. Sometimes the best students in the class are victims of boredom.

A second problem has to do with where and by whom mathematics courses for elementary school teachers are taught. Early in this century most elementary school teachers were trained in two-year normal schools which focused on teaching methods and on matters of professional education. Since that time the responsibility for this training has shifted to four-year colleges and universities. This change, together with the efforts of CUPM (Committee on the Undergraduate Program in Mathematics) in the early 1960's, has resulted in most states requiring that prospective elementary school teachers take at least one mathematics course (three in some states) taught in the mathematics department of a four-year college or a university. What happens there depends in large measure on the history of the institution. Many institutions evolved from normal school to teacher's college to a general college or university. In such an institution the mathematics department has evolved from a service department to the education school and is more likely to have a tradition of serious faculty commitment to the training of elementary school teachers. This commitment is often embodied in tenured faculty with doctoral degrees in mathematics education who teach the courses for elementary school teachers as a regular part of their load. Other institutions with different histories often find themselves with no regular faculty who see these courses as an important professional commitment. The classes are either passed around from semester to semester or are taught by part-time staff or graduate students. Needless to say, there will be a difference between classes taught by a professional who sees the course as a long-term professional commitment and classes taught by others with less commitment.

We find ourselves with a course (or courses) taught in the mathematics department of a four-year college or university by people ranging from mathematics doctoral students, through research mathematicians, to mathematics educators. Moreover, the course has very modest content goals involving arithmetic, elementary geometry, and a smattering of other topics. Yet the course is for college credit!

It is no surprise that a variety of different courses has resulted. One extreme response to the problem of teaching arithmetic for college credit is a Landau-like course on the foundations of arithmetic in which the basic properties of numbers are derived from a limited set of axioms. A more reasonable, and much more widely used, approach is a semiformal course that attempts to attach some meaning to the material. The definitions are fairly precise and axioms are given, but some pictures are drawn, some examples are given, and exercises are included. "Classic" texts embodying this approach include *Theory of Arithmetic* by Peterson and Hashisaki, Wiley (1963), and *Elementary Geometry* by Haag, Hargrove, and Hill, Addison-Wesley (1970). This particular approach is consistent with that of SMSG (School Mathematics Study Group), which was an important force in school mathematics.

Students and educators alike found little relevance in formal arithmetic and geometry for teaching mathematics to children. In response to this perceived need for relevance, a number of textbooks have been written that in one way or another reflect the elementary school curriculum and the problems of teaching it to children. For example, the following quotation appears in the preface to *Elementary Mathematics for Teachers* by John L. Kelley and Donald Richert, Holden-Day, 1970.

> The book as well as the course for which it is written, is somewhat unconventional. Most of the material is organized in the order in which it appears in the elementary curriculum . . . There is considerable emphasis on physical phenomena, on experimentation, and on other forms of nonverbal and nonpictorial teaching. From the viewpoint of the professional mathematician, this course is a mixture of mathematics, science and metaphysics . . . All innovations are based on the conviction that this is not just a course in mathematics, and that the main purpose of the course is to help teachers teach children mathematics.

A number of texts have followed in the tradition of Kelley and Richert. In addition to a less formal and more child-oriented approach, they have included aids such as pages from elementary textbooks, articles from the *Arithmetic Teacher*, and projects to be done by prospective teachers and by children.

The movement toward the elementary school was carried a step further by the Mathematics-Methods Program units by John F. LeBlanc, Donald R. Kerr, Jr., and Maynard Thompson, Addison-Wesley (1976). These materials combined mathematics content and the methods of teaching the content. They also employed an activities format which was designed to involve students in learning by problem-solving or discovery. Another book combining content and methods was published by Richard Lesh, Karen Fuson, and Max Bell, *Algebraic and Arithmetic Structures*: *A Concrete Approach for Elementary School Teachers*, Free Press (1976). The teaching methods emphasis in these materials tends to limit their appeal to courses where mathematics educators are involved.

While the title of the text being reviewed (remember, this is a review!) begins "A Problem Solving Approach to... ," the role of problem solving is different from that in the Mathematics-Methods Program. There is no attempt to present the content of the book using discovery or problem solving; problem solving is instead one of the topics of the book. Each chapter begins with the statement of a challenging problem related to the content of the chapter. At the end of the chapter, the problem, usually a puzzle-type problem, is analyzed under Pólya's headings—understanding the problem, devising a plan, carrying out the plan, and looking back. In addition, the first and last chapters discuss the problem-solving process in some detail. This emphasis on problem solving is consistent with the current attention being given to applications and problem solving by the mathematics education community.

The book is otherwise somewhat less progressive. The presentation of the content is more formal than that in many current competitors. There is also less emphasis on developing meaning and understanding. While the book has positive features, it does not seem to be the quintessential effort to aid in developing competence, confidence, and perspective in prospective elementary school teachers.

It is unlikely that one textbook will be appropriate for the diverse situations in which elementary school teachers are trained. Even so, there is room for improvement in future texts. How can one write better texts? First, pick your content battles and win them! Instead of trying to teach everything, choose a smaller number of topics and then teach them carefully and meaningfully with adequate examples and exercises which develop both skill and understanding. This approach should help reduce the tension between competence and confidence. Second, develop perspective by presenting content in blocks that hang together. For example, in arithmetic show how the properties of our numeration system coordinate with the algorithms so that a knowledge of the basic facts, the numeration system, and the algorithms enable one to perform any computation mechanically. In geometry focus on identifying important shapes in the environment and determining what special properties each shape has that cause it to be used as it is. The important idea in elementary probability is the quantification of chance; that should be emphasized. An important use of statistics is to estimate probabilities based on experience; this would be a nice connection to make.

We hope that efforts will continue to be made to write appropriate books for this difficult market and that these books will be used in courses which receive serious attention from well-trained professionals.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 2 February 1982

Contents

ISSN 0002-9890

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is \$20 included as part of the annual dues of \$40. Students receive a 50% discount. The library subscription price is \$50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

AWARD FOR DISTINGUISHED SERVICE TO DR. THORNTON CARL FRY

G. BALEY PRICE

Department of Mathematics, University of Kansas, Lawrence, KS 66045

The MAA's 1982 Award for Distinguished Service honors Dr. Thornton C. Fry for his many contributions as an industrial mathematician.

Thomas A. Edison established the first industrial research organization at Menlo Park in 1876; in 1878 he employed Francis R. Upton as a mathematical assistant. In 1888 General Electric's Charles Proteus Steinmetz was the only industrial mathematician who was a member of the American Mathematical Society. The Bell Telephone System employed George A. Campbell in 1897, and it developed a staff, during the next several decades, which contained a large proportion of all the industrial mathematicians.

T. C. Fry was born in Findlay, Ohio, on January 7, 1892, and he received his A.B. degree from Findlay College in 1912. In 1913 he received an A.M. degree from the University of Wisconsin, where he was an instructor from 1912 to 1916. In that year Western Electric Company offered him a position in its Engineering Research Department in New York, to complement the engineers and physicists then in it. He accepted. Almost immediately he became prominent in mathematical circles. A list [**1**] of American mathematicians who were engaged in war work and service during World War I contains the following entry: "T. C. Fry, New York City; Industrial Research; War Problems." T. C. Fry was elected a member of the American Mathematical Society on February 23, 1918, and a member of the Mathematical Association of America in September 1921; he has been a member of both organizations ever since. In 1920 he received a Ph.D. degree from Wisconsin; his thesis was a paper he had written at Western Electric in 1918 to resolve a question that was being rather hotly debated at that time: whether the unorthodox uses of nonconvergent series, which the English engineer Oliver Heaviside was making, could be trusted to give correct results.

Dr. Fry made an important contribution to industrial mathematics through the establishment, in May 1922, of a small, separate mathematics section in the Engineering Research Department of Western Electric; in the beginning it consisted of Dr. Fry and a small computing group, commissioned to provide, on a consulting basis, mathematical assistance to any engineer at the location who wished it; soon a program of its own and a budget of its own were added. In 1925, when some parts of A T & T and Western Electric were combined to create Bell Telephone Laboratories, it became a part of BTL. Since then it has grown into the present-day Mathematics and Statistics Research Center of Bell Laboratories, an organization that had 45 mathematicians and statisticians in 1978. As far as we know, it was an innovation in 1922, and even in 1940 Dr. Fry wrote [**2** (a), p. 271]: "So far as is known, mathematicians have not been organized into separate administrative groups in other industries." He was its head until 1944.

Dr. Fry was engaged not only in industrial research but also in teaching and in the writing of textbooks. The notes for one of the "out-of-hour courses" he taught for Western Electric were revised and used in a course of lectures which he gave at the Massachusetts Institute of Technology during the second semester of 1926-1927; they became Dr. Fry's well-known *Probability and Its Engineering Uses*, a book [**4**] published in 1928. This textbook treats, among other topics, the telephone applications of probability theory [**3** (a), p. 924]; the probability problems arising in telephone switching were early examples of the modern theory of queues. Another of the "out-of-hour courses" taught by Dr. Fry led to the publication of his *Elementary Differential Equations* in 1929. Both books received favorable and even enthusiastic reviews in the *AMS Bulletin* and this MONTHLY.

Dr. Fry was a lecturer in mathematics at Princeton University during 1929-1930. In these lectures he applied the methods and results of his thesis [**11**] to discuss the convergence of series of *discontinuous* orthogonal functions. The notes were prepared with publication in mind, but the idea was abandoned.

THORNTON CARL FRY

As World War II approached, the mathematicians called on Dr. Fry to assist in the establishment of *Mathematical Reviews*. A report of the annual meeting, held December 27–30, 1938, states [**6**, p. 203] that Dr. T. C. Fry was one of a committee of six members "appointed to ascertain whether the time is favorable for starting an abstract journal in America under the auspices of the Society... ." Nathan Reingold has recently published an account of the controversy generated by the proposal to publish an abstract journal in the United States [**7**, pp. 327–333]. The committee quickly decided to establish *Mathematical Reviews* [**6**, pp. 641–643], and an "Executive Subcommittee consisting of Oswald Veblen (chairman), Thornton C. Fry, and Warren Weaver was appointed to officiate in setting up the machinery to get the journal under way." Professor Veblen assigned primary responsibility for financing to Weaver; for format, typography, and choice of printer to Fry; and for editorial and personnel matters, to himself.

In 1940 Dr. Fry wrote [**2**] an important report entitled "Industrial Mathematics." The abstract of this report states: "The report consists of three major sections. The first discusses mathematical specialists in industry, calls attention to the essentially consultative character of their work, and makes some observations regarding the education, employment, and supervision of this type of personnel. The second section deals, not with the work of these specialists, but with the uses to which mathematics is put at the hands of industrial workers in general, the various ways in which it contributes to the economy and effectiveness of research, and the kinds of mathematics that are most used. ... The third section is devoted to statistics ... ." The characterization of the industrial mathematician contained in this report is still the accepted one today. The report was important because it described and assessed the state of industrial mathematics in 1940 and set the stage for the spectacular growth of this field that has occurred since World War II. But Dr. Fry underestimated the magnitude of the explosion. Under the heading of "Future Demand" he estimated the number to be "several times 150 a decade or so hence."

In his last published paper, using counting methods throughout, he arrived at the figures 1, 15, 150, and 1,800 in 1888, 1913, 1938, and 1963, respectively (that is, on the birth and on the 25th, 50th, and 75th birthdays of AMS) [**12**].

History records [**1**] that Dr. Fry was engaged in industrial research on war problems in World War I; he was far more deeply and extensively engaged in war work during World War II. In 1939 Warren Weaver was called upon to organize and direct a section of the National Defense Research Committee (later known as OSRD) called the Fire-Control Section [**8**, p. 77], and he selected [**8**, p. 79] three key associates to work with him. They were: "Samuel H. Caldwell, then a professor of electrical engineering at MIT who had been associated with Bush in the development of electrical and mechanical computing devices; Thornton Carl Fry, the head of the mathematics group at the Bell Telephone Laboratories and an extremely clear-headed person with imagination as well as knowledge of the analytical theory and the practical construction of all types of electrical devices, and Edward J. Poitras." (Poitras had most recently been involved in designing the automatic control system for the 200-inch telescope on Mount Palomar.)

One of the first and, as it turned out, one of the most important matters to come before it was a proposal to develop an *electric* gun director to control the fire of anti-aircraft guns. The idea had been conceived by two young engineers at Bell Telephone Laboratories, but had received a poor reception from the Ordnance Department. Dr. Fry brought it before the Fire Control Section, which, after careful consideration, recommended that it be pursued under OSRD sponsorship. The Army concurred [**8**, Chapter 6]. The production model was designated the M-9 gun director [3 (b), Chapter 3]; it was spectacularly successful and made a great contribution to the war effort, especially in the defenses against the V-1 buzz bomb. In the fall of 1942, a reorganization of OSRD created a new OSRD agency called the Applied Mathematics Panel (AMP) [**8**, p. 87]; [**9**, pp. 608-610]. Its members were Richard Courant, G. C. Evans, T. C. Fry, L. M. Graves, Marston Morse, Oswald Veblen, and S. S. Wilks; Warren Weaver and T. C. Fry were its Chairman and Deputy Chief, respectively. A recent article by Mina Rees [**9**] has described the work of the Applied Mathematics Panel. In 1948 Dr. Fry received the Presidential Certificate of Merit for his defense activites.

As a result of war interruptions and dislocations, Dr. Fry's position at Bell Telephone Laboratories changed. He became Director of Switching Research, 1944–1947; Director of Switching Research and Engineering, 1947–1949; Assistant to the Executive Vice President, 1949–1951; Assistant to the President, 1951–1956.

Retirement in 1956 did not end his activites. His biography in *Who's Who in America* records that he held positions as consultant on Research and Development to the International Telephone and Telegraph Corporation, 1956–1957; senior consultant and Vice President in the Univac Division of Sperry Rand Corporation, 1956–1960. From 1961 to 1967 he was consultant to the Director of the National Center for Atmospheric Research in Boulder, Colorado; and he held other consultantships until his final retirement in 1968.

Dr. Fry's interests have been exceptionally broad. He has had some mathematical interest in computing for its own sake, because he patented a mechanism for computing in 1920, and in 1937 he invented the isograph, a mechanical analog machine for finding the complex zeros of polynomials of high degree (for a description and photograph of this machine, see [**2**, p. 282]). The isograph became obsolete after Dr. George R. Stibitz, a member of Dr. Fry's department, invented the general purpose relay digital computer.

The breadth and nature of Dr. Fry's interests are indicated also by his memberships in professional organizations. He is a Fellow of the American Physical Society, AAAS, the Institute of Mathematical Statistics, and the Institute of Electrical and Electronics Engineers; he is a member of the American Mathematical Society, the Mathematical Association of America, the American Astronomical Society, the Society for the Promotion of Engineering Education, the Econometric Society, and Sigma Xi.

Dr. Fry participated widely in all of the activities of the mathematicians. He gave talks frequently at MAA national meetings and at MAA section meetings, and he served on many committees. He was an MAA trustee in 1931 and an MAA vice president in 1936; he was an associate editor of this MONTHLY for at least six years beginning in 1935. He served the American Mathematical Society as executive secretary of its Semi-Centennial observance in 1938. He wrote a paper entitled "A Mathematical Theory of Rational Inference," and participated in research conferences of the economists. Finally, Dr. Fry leaves some challenges. In an address [**10**] at the Ann Arbor meeting in 1955, he described the need to complete the mathematician's education at an earlier age and suggested ways to accomplish this objective. He then wrote: "If these things were done, could we not teach calculus to ninth-grade students? And could we not teach it in the complex plane at the start?"

References

1. D. A. Rothrock, American mathematicians in war service, this MONTHLY, 26 (1919) 40-44.

2. Thornton C. Fry, Industrial Mathematics: (a) Pages 268-288 of the following volume: Research—A National Resource, vol. 2, Industrial Research, December 1940, National Resources Planning Board, U.S. Government Printing Office, 1941; (b) Dr. Fry's article was reprinted as a supplement to this MONTHLY, 48, II (1941)1-38.

3. M. D. Fagen, ed., A History of Engineering and Science in the Bell System: (a) The Early Years (1875-1925), Bell Telephone Laboratories, Inc., 1975; (b) National Service in War and Peace (1925-1975), Bell Telephone Laboratories, Inc., 1978.

4. T. C. Fry, Probability and Its Engineering Uses, D. Van Nostrand, New York, 1928. A second edition was published in 1965.

5. T. C. Fry, Elementary Differential Equations, D. Van Nostrand, New York, 1929.

6. Bull. Amer. Math. Soc., 45 (1939) 197-208, 641-643.

7. Nathan Reingold, Refugee mathematicians in the United States of America, 1933-1941: Reception and reaction, Ann. of Sci., 38 (1981) 313-338.

8. Warren Weaver, Scene of Change, Charles Scribner, New York, 1970.

9. Mina Rees, The mathematical sciences and World War II, this MONTHLY, 87 (1980) 607-621.

10. T. C. Fry, Mathematics as a profession today in industry, this MONTHLY, 63 (1956) 71-80.

11. T. C. Fry, Applications of modern theories of integration to the solution of differential equations, Ann. of Math., 22 (1921) 182-211.

12. T. C. Fry, Mathematics in industry—The first seventy-five years, Science, 143 (1964) 934-938.

GENIUS AND BIOGRAPHERS: THE FICTIONALIZATION OF EVARISTE GALOIS

TONY ROTHMAN

Center for Relativity, University of Texas, Austin, TX 78712

I. Introduction

In Paris, on the obscure morning of May 30, 1832, near a pond not far from the pension Sieur Faultrier, Evariste Galois confronted Pescheux d'Herbinville in a duel to be fought with pistols, and was shot through the stomach. Hours later, lying wounded and alone, Galois was found by a passing peasant. He was taken to the Hospital Cochin where he died the following day in the arms of his brother Alfred, after having refused the services of a priest. Had Galois lived another five months, until October 25, he would have attained the age of twenty-one.

The legend of Evariste Galois, creator of group theory, has fired the imagination of generations of mathematics students. Many of us have experienced the excitement of Freeman Dyson who writes:

> In those days, my head was full of the romantic prose of E. T. Bell's *Men of Mathematics*, a collection of biographies of the great mathematicians. This is a splendid book for a young boy to read (unfortunately, there is not much in it to inspire a girl, with Sonya Kovalevsky allotted only half a chapter), and it has awakened many people of my generation to the beauties of mathematics. The most memorable chapter is called "Genius and Stupidity" and describes the life and death of the French mathematician Galois, who was killed in a duel at the age of twenty.[1]

Dyson goes on to quote Bell's famous description of Galois's last night before the duel:

> All night long he had spent the fleeting hours feverishly dashing off his scientific last will and testament, writing against time to glean a few of the great things in his teeming mind before the death which he saw could overtake him. Time after time he broke off to scribble in the margin "I have not time; I have not time," and passed on to the next frantically scrawled outline. What he wrote in those last desperate hours before the dawn will keep generations of mathematicians busy for hundreds of years. He had found, once and for all, the true solution of a riddle which had tormented mathematicians for centuries: under what conditions can an equation be solved?[2]

This extract is likely the very paragraph which has given the greatest impetus to the Galois legend. As with all legends the truth has become one of many threads in the embroidery. E. T. Bell has embroidered more than most, but he is not alone. James R. Newman, writing in *The World of Mathematics*, notes: "The term *group* was first used in a technical sense by the French mathematician Evariste Galois in 1830. He wrote his brilliant paper on the subject at the age of twenty, the night before he was killed in a stupid duel."[3] From a description in the famed Bullitt archives of mathematics issued by the University of Louisville library, we learn: "Goaded by a 'mignonne' and two of her slattern confederates into a 'duel of honor,' Galois was shot and killed at the age of 20."[4] Leopold Infeld, in his biography of Galois,[5] invokes a conspiracy theory to explain Galois's death: Galois was considered one of the most dangerous republicans in Paris; the government wanted to get rid of him; a female agent provocateur set him up for the duel with d'Herbinville; et cetera. Fred Hoyle, in his *Ten Faces of the Universe*,[6] attempts a partial inversion of the argument: Galois's ability to carry on complex calculations entirely in his head made him appear distant to others; personal animosities arose with his republican friends; they began to think he was not fully for the cause; Galois in their eyes was the agent provocateur; et cetera. All three authors, Bell, Hoyle, and Infeld, invoke a political cause for the duel, with a mysterious coquette just off center.

Tony Rothman received a B.A. in physics from Swarthmore College and is now completing his doctoral research at the Center for Relativity at the University of Texas in Austin. His dissertation concerns nonstandard cosmological models of the early universe. Next year he will be a visiting fellow at Oxford. He is author of a novel, *The World Is Round* (Ballantine Books, 1978), as well as numerous shorter works.

This article is an attempt to sift some of the facts of Galois's life from the embroidery. It will not be an entirely complete account and will assume the reader is familiar with the story, presumably through Bell's version. Because these authors have emphasized the end of Galois's life, I will do so here. As will become apparent, many of the statements just cited are at worst nonsensical, and at best have no basis in the known facts.

Although a number of the documents presented here are, I believe, translated into English for the first time, it should be emphasized that they are not new, just ignored. There is more known about Galois than recent authors admit. It is my hope that some ambitious historian will find the requisite letter in an attic trunk or a newspaper clipping in the Paris archives to unravel the remaining mysteries.

II. Sources

It is not difficult to trace the story of Galois's brief life through its increasingly embellished incarnations. The primary source of information, containing eyewitness accounts and many relevant documents, is the original study of Paul Dupuy, which appeared in 1896.[7] Dupuy was a historian and the Surveillant Général of the Ecole Normale. Bell, Hoyle, and Infeld all cite it as an important reference but never once explicitly quote it. Indeed, Bell acknowledges[8] that his account is based on Dupuy and the documents in Tannery[9] (see below), but it remains unclear how much Bell has read of Dupuy; for while numerous passages are lifted bodily from Dupuy, other important information in the latter is strangely absent. Dupuy's study itself is lacking a number of important letters and documents. Whether Dupuy was unaware of their existence or chose not to publish them I do not know. He also makes a number of minor errors in chronology. In any case, the first lesson is already learned: those who use Dupuy as their sole source of information must make mistakes. Nevertheless, this original biography is much more complete and accurate than the subsequent dilutions and contains more information than a reading of Bell, Hoyle, or Infeld would even suggest. A translation of Dupuy into English should be undertaken.

Some of the documents not found in Dupuy are contained in Tannery's 1908 edition of Galois's papers.[9] All are contained in the definitive 1962 edition of Bourgne and Azra.[10] This volume contains every scrap of paper known to have been written by Galois, an accurate chronology, facsimiles of some of his original manuscripts, and a number of relevant letters by others. When quoting Galois, I have worked exclusively from this edition.

The memoirs of Alexandre Dumas[11] contain a pertinent chapter, and the *Lettres sur les Prisons du Paris* by François Vincent Raspail[12] are the primary source on Galois's months in prison. Some of these letters are quoted by Dupuy and Infeld. Other references will be cited as they appear.

III. Early Life and Louis-le-Grand

I will not dwell at length on the first sixteen years of Galois's life, for they are reported with fair accuracy by Bell. This is not surprising; his account approaches that of a somewhat abridged translation of Dupuy. The divergences will set in later. Thus this section and the next may be taken as a rather condensed review and criticism of Bell. Infeld and Hoyle, who concentrate most of their energies on the duel, will be dealt with at the appropriate time.

Evariste Galois was born on October 25, 1811, not far from Paris in the town of Bourg-la-Reine, France. His father was Nicholas-Gabriel Galois, who was then thirty-six years of age, and his mother was Adelaide-Marie Demante. Both parents were highly intelligent and well educated in the subjects considered important at the time: philosophy, classical literature, and religion. Bell points out that there is no record of any mathematical talent on either side of the family. A more neutral statement should perhaps be made: no record exists in favor of or against any such talent. M. Galois did possess the talent for composing rhymed couplets with which he would amuse neighbors. This harmless activity, as Bell notes, would later cause his undoing. Evariste seems to have inherited some of this ability, participating in the fun at house parties. For the first twelve years of his life, Evariste's mother served as his sole teacher, giving him a solid background in Greek and Latin, as well as passing on her own skepticism toward religion.

In 1815, during The One Hundred Days, M. Galois was elected mayor of Bourg-la-Reine. He had been a supporter of Napoleon and, in fact, had been elected chief of the town's liberal party during Napoleon's first exile. After Waterloo, he had planned to relinquish his post to his predecessor, but the latter had left the country. Galois demanded to be either confirmed or replaced, and in the confusion managed to keep his office. He served the new King faithfully, but from this point on he met increasing resistance from the conservative elements of his town. It is probably safe to say that the younger Galois inherited his liberal ideas from his parents.

On October 6, 1823, Evariste was enrolled in the Lycée of Louis-le-Grand, a famous preparatory school (which still exists) in Paris. Both Robespierre and Victor Hugo had studied there. Louis-le-Grand is where Evariste's troubles began, where Infeld's account of his life essentially opens, and where Bell introduces his theme of "Genius and Stupidity," taking on the tone of a blanket condemnation of almost everyone and everything that surrounded Galois. "Galois was no 'ineffectual angel,'" Bell writes in his introduction, "but even his magnificent powers were shattered before the massed stupidity aligned against him, and he beat out his life fighting one unconquerable fool after another."[13] I believe we will see that the problems ran much deeper than that.

Bell's first liberties with Dupuy are minor. Bell describes Louis-le-Grand as a "dismal horror" and goes on to say "the place looked like a prison and was."[14] Admittedly, Dupuy writes that Louis-le-Grand happened to look like a jail because of its grills, but he then goes on to describe the underlying "passions of work, academic triumph, passions of liberal ideas, passions of memories of the Revolution and the Empire, contempt and hate for the legitimist reaction."[15] Bell, by cutting Dupuy's sentence in half, has begun the slant toward the negative.

At this particular time, there were problems. During Galois's first term, the students, who suspected the new provisor of planning to return the conservative Jesuits to the school, protested by staging a minor rebellion. When required to sing at a chapel service they refused. When required to recite in class, they refused. When required to toast King Louis XVIII at an official school banquet, they refused. The provisor summarily expelled the forty students whom he suspected of leading the insurrection. Galois was not among those expelled, nor is it known if he was even among the rebels, but we may guess that the arbitrariness of the provisor and the general severity of the school's regime made a deep impression on him.

Nevertheless, Galois's first two years at Louis-le-Grand were marked by a number of successes. He received a prize in the General Concourse and three mentions. At this point we witness the first of Bell's distortions of chronology to give the impression that Galois was misunderstood and persecuted. Galois was asked to repeat his third year because of his poor work in rhetoric. Bell writes, "His mathematical genius was already stirring," and "He was forced to lick up the stale leavings which his genius had rejected."[16] I cannot say for certain whether Galois's mathematical genius was already stirring, but it is known that Galois did not enroll in his first mathematics course until *after* he had been demoted.

During this first mathematics course, which he began in February 1827, Galois discovered Legendre's text on geometry, soon followed by Lagrange's original memoirs: *Resolution of Numerical Equations*,* *Theory of Analytic Functions*, and *Lessons on the Calculus of Functions*. Doubtlessly, Galois received his initial ideas on the theory of equations from Lagrange. I do not understand why Bell claims Galois's classwork was mediocre; his instructor, M. Vernier, constantly writes such accolades as "zeal and success," "zeal and progress very marked."[17]

With his discovery of mathematics, Galois became absorbed and neglected his other courses. Before enrolling in M. Vernier's class, typical comments about him had been:[18]

Religious Duties—Good	Work—Sustained
Conduct—Good	Progress—Marked
Disposition—Happy	Character—Good, but singular

*Here and elsewhere, read "algebraic equations" for "numerical equations."

After a trimester in M. Vernier's class, the comments were:

Religious Duties—Good	Work—Inconstant
Conduct—Passable	Progress—Not very satisfactory
Disposition—Happy	Character—Closed and original

The words "singular," "bizarre," "original," and "closed" would appear more and more frequently during the course of Galois's career at Louis-le-Grand. His own family began to think him strange. His rhetoric teachers would term him "dissipated." Bell discusses these remarks at some length. His use of the indefinite pronoun "they" gives the impression that the entire faculty was aligned against Galois. A perusal of Dupuy's appendix, however, shows the negative remarks were penned, by and large, by Galois's two rhetoric teachers. Until this point in his life, I believe it fair to say that Galois was somewhat misunderstood by his teachers in the humanities, but not that he was persecuted.

Slightly more serious problems were soon to arise. His mathematics teacher, M. Vernier, constantly implored Galois to work more systematically. His remark on one of Galois's trimester reports makes this clear: "Intelligence, marked progress, but not enough method,"[19] Galois did not take the advice; he took the entrance examination to l'Ecole Polytechnique a year early, without the usual special course in mathematics, and failed. Apparently he did not know some basics. To Galois, his failure was a complete denial of justice. This and subsequent rejections embittered him for life. When we examine some of his later writings, I think it will be evident that he developed not a little paranoia.

Galois did not yet give up. The same year, 1828, saw him enroll in the course of Louis-Paul-Emile Richard, a distinguished instructor of mathematics. Richard encouraged Galois immensely, even proclaimed that he should be admitted to the Polytechnique without examination. The results of such encouragement were spectacular. In April of 1829, Galois published his first small paper, "Proof of a Theorem on Periodic Continued Fractions." It appeared in the *Annales de Gergonne*.

This paper was a minor aside. Galois had also been working on the theory of equations ("Galois theory"). On May 25 and June 1, 1829, while still only 17, he submitted to the Academy his first researches on the solubility of equations of prime degree. Cauchy was appointed referee.

We now encounter a major myth which evidently has its origins in the very first writings on Galois and which has been perpetuated by virtually all writers since. This myth is the assertion that Cauchy either forgot or lost the papers (Dupuy, Bell[20]) or intentionally threw them out (Infeld[21]). Recently, however, René Taton has discovered a letter of Cauchy in the Academy archives which conclusively proves that he did not lose Galois's memoirs but had planned to present them to the Academy in January 1830.[22] There is even some evidence that Cauchy encouraged Galois. The letter and related events will be discussed in more detail later; for now we note only that to hold Cauchy responsible for "one of the major disasters in the history of mathematics," to paraphrase Bell,[23] is simply incorrect, and to add neglect by the Academy to the list of Galois's difficulties during this period appears entirely unwarranted.

A truly tragic blow came within a month of the submissions: on the second of July, 1829, Galois's father committed suicide. The reactionary priest of Bourg-la-Reine had signed Mayor Galois's name to a number of maliciously forged epigrams directed at Galois's own relatives. A scandal erupted. M. Galois's good nature could not withstand such an attack and he asphyxiated himself in his Paris apartment "not two steps from Louis-le-Grand." During the funeral, when the same clergyman attempted to participate, a small riot erupted. The loss of his father may explain much of Galois's future behavior. We must wait a few years, until Evariste's second prison term, to see this. In any case, he loved his father dearly, and if an iron link had not already been forged between the Bourbon government and the Jesuits, it had now.*

*Probably the clearest picture of the relationship between the Jesuits and the Bourbons, one which contains episodes paralleling that of M. Galois's misfortunes, is Stendhal's famous novel *The Red and The Black*.

But Galois's troubles were not yet over. A few days later, he failed his examination to l'Ecole Polytechnique for the second and final time. Legend has it that Galois, who worked almost entirely in his head and who was poor at presenting his ideas verbally, became so enraged at the stupidity of the examiner that he hurled an eraser at him. Bell records this as a fact,[24] although Dupuy specifically states that it is only an unverified tradition.[25] The examination failures, as well as the misunderstanding of his humanities teachers, left him irrevocably embittered. Bell quotes him as writing, "Genius is condemned by a malicious social organization to an eternal denial of justice in favor of fawning mediocrity."[26] I believe Bell constructed this quotation from a passage of Dupuy[27] but Galois did express similar sentiments in his fragmentary essay "Sciences Hiérarchie: Ecoles" and in "Sur l'Enseignment des Sciences" ("Hierarchy is a means for the inferior."[28]). In Bell's diatribe against this famous examination, as well as in other accounts of it, the suggestion that the death of Galois's father several days before may have had something to do with the outcome never arises. It is a simple matter for Bell to lay the fault squarely with the examiner's stupidity because Bell has placed the examination before M. Galois's unfortunate suicide. In this case, Bell is not fully to blame; Dupuy does not date the examination.[29] I do not wish to suggest Galois should have been failed. I only wish to point out that the examination must have been held under the worst possible conditions.

Thus, Galois's secondary school career ended in a string of minor setbacks and two major disasters. Evariste had not planned to take the Baccalaureate examinations, because the Ecole Polytechnique did not require them. Now, having failed the Polytechnique's entrance examination and having decided to enter the less prestigious Ecole Normale,* he was forced to reconsider. "Still persecuted and maliciously misunderstood by his preceptors," in Bell's words, "Galois prepared himself for the final examinations."[30] Despite such malice, Galois did very well in mathematics and physics, although less well in literature. He received both a Bachelor of Letters and a Bachelor of Science on the twenty-ninth of December, 1829.

It is interesting to note that, although he has continued to play the role of muckraker of malice, Bell has failed to mention M. Richard's distinct cooling toward Galois, on whom he had previously bestowed encomia. After the first trimester of the 1828–1829 academic year, Richard wrote: "This student is markedly superior to all his classmates." After the second: "This student works only in the highest realms of Mathematics." After the third: "Conduct good, work satisfactory."

Because I do not have an accurate date for this report, I cannot propose a specific event as the cause of this obvious change in attitude. Presumably, it occurred in the spring of 1829, shortly before or after Galois's time of troubles began. One could, of course, argue that M. Richard had simply become bored with Galois. Otherwise, it does serve to show that Bell's black-and-white presentation of Galois's preceptors is an oversimplification.

IV. L'Ecole Normale

The early months of 1830, which saw Galois officially enrolled as a student at l'Ecole Normale, also witnessed an interesting series of transactions with the Academy. As will be recalled, Galois submitted his first researches to the Academy on May 25 and June 1 of 1829. On January 18, 1830, Cauchy wrote the previously mentioned letter discovered by Taton:[31]

> I was supposed to present today to the Academy first a report on the work of the young Galois, and second a memoir on the analytic determination of primitive roots in which I show how one can reduce this determination to the solution of numerical equations of which all roots are positive integers. Am indisposed at home. I regret not to be able to attend today's session, and I would like you to schedule me for the following session for the two indicated subjects.
>
> Please accept my homage . . .
>
> A.-L. Cauchy

*During the restoration, l'Ecole Normale was actually called l'Ecole Préparatoire.

This letter makes it clear that, six months after their receipt, Cauchy was still in possession of Galois's manuscripts, had read them, and very likely was aware of their importance. At the following session on 25 January, however, Cauchy, while presenting his own memoir mentioned above, did not present Galois's work. Taton hypothesizes that between January 18 and January 25, Cauchy persuaded Galois to combine his researches into a single memoir to be submitted for the Grand Prize in Mathematics, for which the deadline was March 1. Whether or not Cauchy actually made the suggestion cannot yet be proved, but in February Galois did submit such an entry to Fourier in his capacity as perpetual secretary of mathematics and physics for the Academy. In any case, there is an additional piece of evidence which attests to Cauchy's appreciation of Galois's work. This is an article which appeared the following year on 15 June 1831 in the Saint-Simonian journal *Le Globe*. The occasion was an appeal for Galois's acquittal after his arrest following the celebrated banquet at the *Vendanges des Bourgogne*:

> Last year before March 1, M. Galois gave to the secretary of the Institute a memoir on the solution of numerical equations. This memoir should have been entered in the competition for the Grand Prize in Mathematics. It deserved the prize, for it could resolve some difficulties that Lagrange had failed to do. *M. Cauchy had conferred the highest praise on the author about this subject.* And what happened? The memoir is lost and the prize is given without the participation of the young *savant*. [Taton's italics.][32]*

The misfortune referred to above was the death of Fourier in April. Galois's entry could not be found among Fourier's papers. In Galois's eyes this could not be an accident. "The loss of my memoir is a very simple matter," he wrote. "It was with M. Fourier, who was supposed to have read it and, at the death of this *savant*, the memoir was lost."[33] It was an unfortunate coincidence; however it was not Fourier's sole responsibility to read the manuscript, for the committee appointed to judge the Grand Prize consisted also of Lacroix, Poisson, Legendre, and Poinsot.[34] I mention this because a number of sources give the impression that somehow Fourier either intentionally lost the paper or could not understand it.[35]

In spite of the setback caused by the loss of his manuscript, April saw the publication of Galois's paper "An analysis of a Memoir on the Algebraic Resolution of Equations" in the *Bulletin de Ferussac*. In June he published "Notes on the Resolution of Numerical Equations" and the important article "On the Theory of Numbers."[36]

In addition to propagating the legend that Cauchy lost the manuscripts, Bell, curiously, does not mention Fourier by name in the preceding misadventure, although Dupuy is explicit on the identity of the Academy's Perpetual Secretary. I do not understand the reason for this omission unless Bell felt it a little too much to "expose" Cauchy, Fourier, and later Poisson, as incompetents. Bell also does not make it clear that the papers listed above (plus a later memoir) constitute what is now called Galois theory. If this point had been clarified, the claim that Galois had written the theory down on the eve of the duel would be difficult to substantiate or even to suggest.

From this point onward, the scenario of Galois as a passive victim of negligence, misunderstanding, and bad luck begins to break down—if it has not already. More and more he participated in the creation of his own disasters. But this picture does not fit Bell's plan. Therefore chronology is rearranged, events are omitted, and others invented in increasing quantity, until the end of his account is largely fantasy. The wholesale reordering of events will be especially evident in what follows.

---

*My own interpretation of this article is slightly different from that of Taton. Taton writes that the journalist evidently had first-hand information. But note the date: 15 June 1831. In the aftermath of the July revolution, Cauchy fled France during September, almost nine months prior to the article's publication. It is difficult to see when the journalist would have spoken to Cauchy. However, the article appeared in a Saint-Simonian journal. Galois's best friend, Auguste Chevalier, was one of the most active Saint-Simonians. My own suspicion, which I cannot prove, is that the journalist was Chevalier and the information was coming directly from Galois. If this hypothesis is correct, Galois himself is admitting Cauchy's encouragement.

Most important, Bell gives an extremely late start to Galois's political activities. He remarks that had Evariste's teachers at Louis-le-Grand allowed him to study only mathematics he might have lived to be eighty.[37] Unlikely. According to Dupuy, one of the reasons Galois had hoped to attend the Polytechnique was to participate in political activities. At l'Ecole Normale he became a "polytechnician in exile." The July revolution of 1830 reared its head. The Director of l'Ecole Normale, M. Guigniault, locked the students in so that they would not be able to fight on the streets. Galois was so incensed at the decision that he tried to escape by scaling the walls. He failed, and in doing so missed the revolution. Afterwards, the Director put the students in the service of the provisional government. Charles X had fled France. He would be followed in September by Cauchy. Louis-Phillipe was the new King.

The events of July, severely abridged here, Bell chronicles accurately. He does fail to mention that Galois probably joined the Society of the Friends of the People, one of the most extreme republican secret societies, within the next month, certainly before December.[38] The importance of this omission will be explained after we have filled in the remaining gaps of the narrative.

In December of that year, M. Guigniault was engaging in polemics against students in the pages of several newspapers. Galois saw his chance for attack and jumped into the squabble with a blistering letter to the *Gazette des Ecoles*. It read in part:

> Gentlemen:
>
> The letter which M. Guigniault inserted yesterday in the *Lycée* on the occasion of one of the articles in your journal has seemed to me very inappropriate. I had thought that you would welcome with eagerness every means to expose this man.
>
> Here are the facts which can be verified by forty-six students.
>
> On the morning of July 28, when many of the students wished to leave the school and fight, M. Guigniault told them on two occasions that he would call the police to reestablish order within the school. The police on the 28th of July!
>
> On the same day, M. Guigniault told us with his usual pedantry: "There are many brave men fighting on both sides. If I were a soldier I would not know what to decide—to sacrifice liberty or LEGITIMACY?"
>
> Here is the man who the next day covered his hat with an immense tricolor cockade. Here are our liberal doctrines![39]

Galois continues. According to Dupuy, every statement in the letter is accurate. Nonetheless, the result was what might have been anticipated: Galois was expelled. The action was to become official on January 4, but Galois quit school immediately and joined the Artillery of the National Guard, a branch of the militia which was almost entirely composed of republicans. It is interesting that the forty-six students referred to in the letter actually published a reply *against* Galois, but this seems to have been at the "prompting" of M. Guigniault.[40]

December was a turbulent month for other reasons. After the Bourbons had fled France, four of their ex-ministers were tried for treason. Popular sentiment called for their execution. The decision to execute or imprison for life was to be announced on December 21. That day, the Artillery of the National Guard was stationed in the quadrangle at the Louvre. Galois was certainly there. The atmosphere was very tense. If the ministers were given a life sentence, the artillerymen had planned to revolt. But the Louvre was soon surrounded by the full National Guard and troops of the line, more trustworthy arms of the military. A distant cannon shot was heard. It signaled the end of the trial and that the ministers had indeed been given imprisonment over execution. The artillerymen and the National Guard readied themselves for bloodshed, but with the arrival at the Louvre of thousands of Parisians, the fighting did not erupt. Over the next few days, the situation in Paris grew calmer with the appearance of Lafayette, who called for peace, and daily proclamations calling for order. On December 31, 1830, the Artillery of the National Guard was abolished by royal decree in fear of its threat to the throne.[41]

In January 1831 Galois, no longer a student, attempted to organize a private class in mathematics. At the first meeting, about forty students appeared[42] but the endeavor did not last long, evidently because of Galois's political activities. On the 17th of that month, upon the

invitation of Poisson, Galois submitted a third version of his memoir to the Academy. Later, in July, Poisson would reject the manuscript.* This rejection will be discussed at the proper time, but we should note that by that time Galois would have already been arrested.

If we return to Bell's account, we will now find a totally distorted chain of events: The months after July are missing; Galois still has not joined the Society of Friends of the People. He leaves school in December but has not joined the artillery. The events at the Louvre, which will turn out to have critical importance for the remainder of the story, never take place. Galois attempts to organize his private course in mathematics. Bell writes: "Here he was at nineteen, a creative mathematician of the first rank, peddling to no takers... Finding no students, Galois temporarily abandoned mathematics and joined the Artillery of the National Guard..."[43] According to Bell, Galois submits his paper to Poisson, it is rejected; and this being the "last straw," Galois decides to devote "all his energy to revolutionary politics."

The chronology presented by Bell is thus completely backwards. The impression given by this rearrangement of events is once again that of a misunderstood and persecuted Galois who, surrounded on all sides by idiots, finally gives up and goes into radical politics. By writing that Galois found no students, Bell of course strengthens this impression. A more balanced account clearly requires what is lacking in Bell: a Galois of volition. We may get a better indication of his character and behavior during the spring of 1831 from a letter written on April 18 by the mathematician Sophie Germain to her colleague Libri:[44]

> ...Decidedly there is a misfortune concerning all that touches upon mathematics. Your preoccupation, that of Cauchy, the death of M. Fourier, have been the final blow for this student Galois who, in spite of his impertinence, showed signs of a clever disposition. All this has done so much that he has been expelled from l'Ecole Normale. He is without money and his mother has very little also. Having returned home, he continued his habit of insult, a sample of which he gave you after your best lecture at the Academy. The poor woman fled her house, leaving just enough for her son to live on, and has been forced to place herself as a companion in order to make ends meet. They say he will go completely mad, I fear this is true.

Unfortunately, as Bell observes, Galois was no ineffectual angel.

Before continuing, another historical detail should be mentioned. As an aftermath of the December events at the Louvre and the dissolution of the Artillery of the National Guard, nineteen officers were arrested, having been suspected of planning to deliver their cannons to the people. The charge was conspiracy to overthrow the government. In April, all nineteen were acquitted.

Until now, my criticism has been devoted almost entirely to Bell. Partly, this has been because his account is by far the most famous. There are other reasons as well. Hoyle's short essay, as already mentioned, is purely concerned with Galois's death and thus has little to say concerning the foregoing events. Infeld's account, on the other hand, is of book length. In a single article it would be difficult to debate all salient points. Nonetheless, Infeld has also stated[45] that he is primarily concerned with the events surrounding the duel. It is then reasonable to devote attention here to that aspect of the book. Infeld's work is actually something of a curiosity. The bulk of it is a fictionalized biography, interspersed with real documents and eyewitness accounts. All dates, names, and places are respected. The second part of the biography consists of a lengthy Afterword in which Infeld details exactly what he has invented, what he has not, and what he believes to be true. He also includes a fairly comprehensive bibliography. In my criticisms of Infeld to follow, I only take issue with those points he claims not to have invented. The reader may get the flavor of the author's intent by noting that at Galois's private class in algebra, spoken of earlier, Infeld has stationed two police spies.[46]

It might also be noted that, according to James R. Newman's brief remark quoted in the Introduction, Galois at this point in the narrative would be dead.

---

*Dupuy does not date this event and its placement in the narrative may have misled Bell.

V. Arrest and Prison

And thus we arrive on May 9, 1831. The occasion was the republican banquet at the restaurant *Vendanges des Bourgogne*, where approximately two hundred republicans were gathered to celebrate the acquittal of the nineteen republicans on conspiracy charges. As Dumas says in his memoirs, "It would be difficult to find in all Paris two hundred persons more hostile to the government than those to be found reunited at five o'clock in the afternoon in the long hall on the ground floor above the garden."[47] It is worth quoting Bell's description of this event:

> The ninth of May, 1831, marked the beginning of the end. About two hundred young republicans held a banquet to protest against the royal order disbanding the artillery which Galois had joined. Toasts were drunk to the Revolutions of 1789 and 1793, to Robespierre, and to the Revolution of 1830. The whole atmosphere of the gathering was revolutionary and defiant. Galois rose to propose a toast, his glass in one hand, his open pocket knife in the other. "To Louis-Phillipe"—the King. His companions misunderstood the purpose of the toast and whistled him down. Then they saw the open knife. Interpreting this as a threat against the life of the King, they howled their approval. A friend of Galois, seeing the great Alexander Dumas and other notables passing by the open windows, implored Galois to sit down, but the uproar continued. Galois was the hero of the moment, and the artillerists adjourned to the street to celebrate their exuberance by dancing all night. The following day Galois was arrested at his mother's house and thrown into the prison at Sainte-Pélagie.[48]

Dumas himself describes this event at length in his memoirs. To save space, we quote only a portion:

> Suddenly, in the midst of a private conversation which I was carrying on with the person on my left, the name Louis-Phillipe, followed by five or six whistles, caught my ear. I turned around. One of the most animated scenes was taking place fifteen or twenty seats from me.
>
> A young man who had raised his glass and held an open dagger* in the same hand was trying to make himself heard. He was Evariste Galois, since killed by Pescheux d'Herbinville, a charming young man who made silk-paper cartridges which he would tie up with silk ribbons.†
>
> Evariste Galois was scarcely 23 or 24 at the time. He was one of the most ardent republicans. The noise was such that the very reason for this noise had become incomprehensible.
>
> All I could perceive was that there was a threat and that the name of Louis-Phillipe had been mentioned; the intention was made clear by the open knife.
>
> This went way beyond the limits of my republican opinions. I yielded to the pressure from my neighbor on the left who, as one of the King's comedians, didn't care to be compromised, and we jumped from the window sill into the garden.
>
> I went home somewhat worried. It was clear this episode would have its consequences. Indeed, two or three days later, Evariste Galois was arrested.[49]

The amusing discrepancies between the two accounts are not entirely difficult to explain. Bell has taken his description from Dupuy, almost word for word, who in turn has based his account on Dumas and the report in the *Gazette des Ecoles*.[50] The toasts Bell mentions as well as the description of the general atmosphere are found in Dupuy. But Bell has mistranslated: Dupuy writes that " ...*Dumas et quelques autres passaient par le fenêtre dans le jardin pour ne pas se compromettre*..."[51] which, in this context, means " ...Dumas and several others jumped through the window into the garden in order not to be compromised." It does not here mean, "Dumas and several others passed by the window in order not to be compromised." One of course wonders how Bell interpreted the clause "in order not to be compromised" in light of his own translation. Why should Dumas be passing by open windows in order not to be compromised? It is difficult to call this carelessness. Bell has also distorted the reason for the banquet. Dupuy clearly states[52] that

*Literally, *poignard*.

†This is a literal translation of Dumas. We have not been able to discover exactly to what this occupation refers, but it is a plausible guess that d'Herbinville made what the British call "crackers" (French, *diablotins*), party favors that pop when the ribbons are pulled and contain inspirational messages. They seem to have been invented at about this time.

it was a celebration for the acquittal of the nineteen conspirators. But Bell has not mentioned the trial. For consistency's sake, he must therefore emphasize the obviously revolutionary character of the gathering.

The issue of accuracy becomes more important when we question the most glaring omission in Bell's account: the absence of any mention of Pescheux d'Herbinville. The single sentence in Dumas is the only extant evidence that d'Herbinville was the man who eventually shot Galois. Although Dumas is repeatedly cited by Dupuy, Bell has obviously not read Dumas. If he had, Bell might have seen fit to close the discrepancies in the banquet accounts, in order not to be compromised. On the other hand, Bell claims to have read Dupuy; Dupuy, once again citing Dumas, explicitly names d'Herbinville as Galois's adversary.[53] Hoyle is guilty of the same charge; listing Dupuy as a main reference, he relegates d'Herbinville to the ranks of anonymous assassins. Infeld, who does identify d'Herbinville, attempts to prove he was a police agent.

For mathematics, of course, it is not important to know exactly who killed Galois; for historical accuracy, it is. In light of the plethora of theories which have arisen to explain the cause of the celebrated duel, most of which involve police spies, agents provocateurs, and political overtones, the identity of d'Herbinville might be a key piece of information. We will, in fact, find that the only evidence strongly indicates d'Herbinville was *not* a police agent. d'Herbinville and the conspiracy theories will be discussed in greater detail later, but for now let us return to Galois.

Evariste was arrested at his mother's house the day following the banquet, which does indicate that police or informers were at the dinner, although the celebration was open to any subscriber. Galois was held in detention at Sainte-Pélagie prison until June 15, when he was tried for threatening the King's life. Bell's description of this event is highly oversimplified. Indeed, the defense lawyer did claim Galois has actually said, "To Louis-Phillipe, *if he betrays*," but that the noise had been such to drown out the qualifying clause. Nonetheless, the matter took on a less facetious aspect when the prosecutor asked Galois if he really intended to kill the King. Galois replied, "Yes, if he betrays." The prosecutor goes on to ask how Galois "can believe this abandonment of legality on the part of the King," and Galois answers, "Everything makes us believe he will soon turn traitor if he has not done so already." Galois is asked to clarify his remarks and basically repeats what he has already said: "I will say that the trend in government can make one suppose that Louis-Phillipe will betray one day if he hasn't already."

As Dumas aptly remarks, "One understands that with such lucidity in the questions and answers, the discussion did not last long." Apparently moved by Galois's youth, the jury acquitted him within moments. Dumas writes, "I repeat that this is a rude generation, perhaps a bit foolish, but you will recall Beranger's song *Les Fous* ["The Fools" or "The Madmen"]."[54]

Shortly after this event, the Academy rejected Galois's memoir on the resolution of equations, this time with Poisson as referee. The rejection was written on July 4, although according to Infeld[55] Galois did not receive the letter until October, when he was in prison again.* By this time, about eight months had passed since he had submitted the paper at Poisson's request. As we will see, Galois did not take the rejection lightly.

The cause for Galois's second arrest was preventive: On Bastille Day, July 14, 1831, he and his republican friend Duchatelet were apprehended dressed in Artillery Guard uniforms and heavily armed. Because the Artillery Guard had been disbanded on the last day of 1830 in fear of its becoming an instrument of the republicans, to wear the uniform was an outright gesture of defiance. It was also illegal. This was the charge brought against Galois, but not until the late date of October 23; he was sentenced to six months in prison. The sentence was confirmed by the court of appeals on December 3. In the meantime, Galois had been sitting in Sainte-Pélagie prison since his arrest in July.

In Bell's fierce diatribe against this arrest he forgets to mention several relevant points. First, he does not seem to comprehend that this was not the Paris of our day but Paris one year after a

*I have found no other source which either corroborates or contradicts Infeld's claim that the rejected manuscript was not received until October, three months after the actual rejection.

revolution, when street riots were rampant, assassination attempts not uncommon, and republican activity dangerous.[56] The "celebration" Bell mentions was a republican demonstration on Bastille Day. Today such a demonstration would be considered patriotic; then it was seditious. This is exactly what the police chief decided when he went on record opposing the demonstration.[57] Bell concedes, "True, Galois was armed to the teeth when arrested, but he had not resisted arrest."[58] More precisely, Galois was carrying a loaded rifle, several pistols, and his dagger, a punishable offense even in our more moderate times. To say that he had not resisted arrest may also be inaccurate. The police came to Galois's house to detain him, but Evariste had already decamped.

Galois's predicament was not helped by his friend Duchatelet, who drew a picture on the wall of his cell of the King's head lying next to a guillotine with the inscription "Phillipe will carry his head to your altar, O Liberty!"[59] Part of the delay in bringing Galois to trial was the fact that Duchatelet was tried first.

The point here is not to argue for or against the justice of Galois's arrest. The point is that he was behaving dangerously in a dangerous time. Two forces are clearly at work here: the government's intention to deal harshly with him after his threat of regicide and his own inability to keep out of trouble.

During his stay in prison, a number of events occurred which throw further light on Galois's personality. These incidents were recorded by the republican François Vincent Raspail. Raspail was an early botanist, one of the first to advocate the use of the microscope to examine cell structure in plants. He also had his troubles with the Academy and was sitting next to Dumas at the May 9 banquet. An ardent republican, he refused to receive the Cross of the Legion of Honor from Louis-Phillipe and during the years 1830–1836 spent a total of twenty-seven months in prison.[60] Later in life, Raspail became a famous statesman. He is now remembered by a boulevard and a metro stop in Paris. He lived to be about eighty. One of his many arrests occurred at about the same time Galois was taken. Raspail recorded the following incidents in several of his letters. Infeld quotes him several times at great length but never explains who he was.

On July 25, 1831, Raspail wrote that his fellow prisoners had taunted Galois into drinking some liquor, a pastime at which he was apparently a novice:

> To refuse the challenge would be an act of cowardice. And our poor Bacchus has so much courage in his frail body that he would give his life for the hundredth part of the smallest good deed. He grasps the little glass like Socrates courageously taking the hemlock; he swallows it at one gulp, not without blinking and making a wry face. A second glass is not harder to empty than the first, and then the third. The beginner loses his equilibrium. Triumph! Homage to the Bacchus of the jail! You have intoxicated an ingenuous soul, who holds wine in horror.[61]

The scene repeats itself. This time Galois empties a bottle of brandy in a single draught. Galois, drunk, pours out his soul to Raspail in haunting prophecy:

> How I like you, at this moment more than ever. You do not get drunk, you are serious and a friend of the poor. But what is happening to my body? I have two men inside me, and unfortunately I can guess which is going to overcome the other. I am too impatient to get to the goal. The passions of my age are all imbued with impatience. Even virtue has that vice with us. See here! I do not like liquor. At a word I drink it, holding my nose, and get drunk. I do not like women and it seems to me that I could only love a Tarpeia or a Graccha.* And I tell you, I will die in a duel on the occasion of some *coquette de bas étage*. Why? Because she will invite me to avenge her honor which another has compromised.
>
> Do you know what I lack, my friend? I confide it only to you: it is someone whom I can love and love only in spirit. I have lost my father and no one has ever replaced him, do you hear me...?[62]

---

*The *Encyclopedia Britannica* states that Tarpeia, according to Roman legend, was the daughter of the Roman Commander in charge of defending the capital against the Sabines. She offered to betray the citadel in exchange for what the Sabines wore on their left arms, i.e., their bracelets. Taking her at her word, the Sabines crushed her under their shields. Graccha refers to Cornelia Graccha, the mother of Tiberius and Gaius, who is remembered as their educator as well as an accomplished author in her own right. Although hostile propaganda later suggested she encouraged her sons' more revolutionary policies, she seems rather to have restrained them.

The aftermath of this episode is neither heartwarming nor pleasant; Galois in a delirium attempts suicide:

> We laid him out on one of our beds. But the fever of intoxication tormented our unhappy friend.... He would fall back senseless only to raise himself with new exaltation, and he foretold sublime things which a certain reserve often rendered ridiculous.
>
> "You despise me, you who are my friend! You are right, but I who committed such a crime must kill myself!"
>
> And he would have done it if we had not flung ourselves on him, for he had a weapon in his hands...[63]

Several important points need to be made about these passages. Bell, in his account, says only, "Goaded beyond endurance, Galois seized a bottle of brandy, not knowing or caring what it was, and drank it down. A decent fellow prisoner took care of him until he recovered."[64] Thus, the really important parts of the episode, which tell us something about Galois's character and which bear on future events, are omitted altogether.

Later, in attempting to understand the cause of Galois's death, Dupuy remarks, "If I credit an allusion of Raspail, Galois lost his virgin heart to *quelque coquette de bas étage*."[65] Bell writes: "Some worthless girl ["*quelque coquette de bas étage*"] initiated him."[66] Here, Bell is taking a conjecture of Dupuy based on a letter of Raspail reporting an utterance of Galois spoken in a delirium a year before the duel as a characterization of real events. This can only be termed fabrication. And it is very likely that this piece of fabrication is responsible for the widespread belief that a prostitute was the cause of Galois's death.

Infeld, in his version of the prison scene, quotes the letters far more fully than Dupuy, but jumps from "Tarpeia and Graccha" to "Do you know what I lack, my friend?" In other words, he omits Galois's prophecy that he will die in a duel. He also makes no comment whatsoever on Galois's suicide attempt. This selective presentation and slanting of evidence is characteristic of Infeld's book. He publishes any document or any portion of a document which does not interfere with his stated hypothesis that Galois was killed by the secret police. More obvious examples will be presented later when we discuss the actual circumstances surrounding the duel.

On August 2, Raspail chronicles an interesting series of events which took place after his previous letters. On July 27, the prisoners were invited to attend a mass in memory of those killed during the July revolution a year earlier. Because many of the prisoners were political, the atmosphere was tense and an open riot was expected to erupt at any moment. A few prudent prison leaders defused the situation, and two days passed without violence. At lock-up time on the 29th, a shot was heard throughout the prison, followed by cries of "Help, murder!" The following day, the mystery was clarified. Raspail, quoting the conversation of another prisoner with the prison superintendent, writes:

> "Here are the facts. I am one of those in the attic room of the bathing pavilion. We were quietly going to bed. The man whose bed is between two casements had his face toward the window while undressing and he was humming a tune."
>
> "At that moment a shot was fired from the garret opposite. We thought our comrade was dead, but he was only unconscious. Not knowing where the shot had come from, nor how serious the wound was, we called for help. For in such a room, open in all directions through six windows, a better-aimed shot would have struck down its man."[67]

The shot, it turned out, came from a garret, across the street, where one of the prison guards lived. Galois was not the man who was at the window and wounded. However, he was in the same room and was later thrown into the dungeon, evidently because he had insulted the superintendent, probably accusing him of having intentionally arranged the shooting. Raspail continues to record the conversation. The prisoner already quoted is talking:

> "What? You have no order to seize the guilty man [the guard who fired the shot]? But you have one to throw into the dungeon both the victim of this shameful trap and the witnesses of it? It may sound insolent to say that the administration pays turnkeys to murder prisoners. But what if this insolent statement is true? And I bear witness that no other insolence has come from those who were thrown into the dungeon. This

young Galois doesn't raise his voice, as you well know; he remains as cold as his mathematics when he talks to you."

"Galois in the dungeon!" repeats the crowd. "Oh, the bastards! They have a grudge against our little scholar."

"Of course they have a grudge against him. They trick him like vipers. They entice him into every imaginable trap. And then, too, they want an uprising."[68]

An uprising they got. This oblique conversation ends with the superintendent taking to his heels as the prisoners take control of the prison. The situation remains stalemated until late that night when the infantry is called in. The prisoners surrender without violence and remarkably no one is hurt.

I have tried to present this episode in as neutral a tone as possible. Infeld interprets the shot as an assassination attempt on Galois's life, and later cites it in his Afterword as his first piece of evidence that Galois was murdered by the government.[69] It is agreed that the moderate government of Louis-Phillipe would have liked to have been rid of all political extremists. But a conspiracy theory presumes there exists a reason to single out a particular victim. Why Galois over Raspail? A shot was fired in a prison full of political prisoners on the verge of a riot, at night ("lock-up time"), into a room containing an unknown number of men, evidently "aimed" at someone else. Yes, it could have been an attempt to kill Galois. I do not find the evidence compelling.

More compelling is the evidence for the absolute hatred Galois had developed for the Academy, which I feel can only be termed paranoid. And, as is not uncommon with paranoiacs, there was a kernel of justification for the behavior. At some point in October, according to Infeld, Galois was notified of Poisson's rejection of his latest manuscript on the theory of equations.* Infeld quotes the following letter, which was originally published by Bertrand. It is not quoted by Dupuy.

Dear M. Galois:

Your paper was sent to M. Poisson to referee. He has returned it with his report, which we quote:

"We have made every effort to understand M. Galois's proofs. His argument is neither sufficiently clear nor sufficiently developed to allow us to judge its rigor; it is not even possible for us to give an idea of this paper.

The author claims that the propositions contained in his manuscript are a part of a general theory which has rich application. Often different parts of a theory clarify each other and can be more easily understood when taken together than when taken in isolation. One should rather wait to form a more definite opinion, therefore, until the author publishes a more complete account of his work."

For this reason, we are returning your manuscript in the hope that you will find M. Poisson's remarks useful in your future work.

François Arago, Secretary to the Academy[70]

Bell, elaborating from Dupuy, states that Poisson found the manuscript "incomprehensible" but "did not state how long it had taken him to reach this remarkable conclusion."[71] I believe this is an unfair characterization of Poisson's comments. This is the rejection that Bell has occurring before Galois's arrest.

In light of previous events and in light of his character, it is not terribly surprising that Galois reacted violently to what might nowadays be considered an encouraging rejection letter. He gave up all plans to publish his papers through the Academy and decided to publish them privately with the help of his friend Auguste Chevalier. Galois collected his manuscripts and in December, while still in Sainte-Pélagie, penned what surely must be one of the most remarkable documents in the history of mathematics, his *Préface*. The entire *Préface* runs about five pages. Infeld, to his credit, prints some of it, although he alters and omits certain parts at will. To save space, I here quote only the first page. The full text can be found in Bourgne and Azra:

*See footnote to page 93.

Firstly, you will notice the second page of this work is not encumbered by surnames, Christian names or titles. Absent are eulogies to some prince whose purse would have opened at the smoke of incense, threatening to close once the incense holder was empty. Neither will you see, in characters three times as high as those in the text, homage respectfully paid to some high-ranking official in science, or to some savant-protector, a thing thought to be indispensable (I should say inevitable) for someone wishing to write at twenty. I tell no one that I owe anything of value in my work to his advice or encouragement. I do not say so because it would be a lie. If I addressed anything to the important men of science or of the world (and I grant the distinction between the two at times is imperceptible) I swear it would not be thanks. I owe to important men the fact that the first of these papers is appearing so late. I owe to other important men that the whole thing was written in prison, a place, you will agree, hardly suited for meditation, and where I have been dumbfounded at my own listlessness in keeping my mouth shut at my stupid, spiteful critics: and I think that I can say "spiteful critics" in all modesty because my adversaries are so low in my esteem. The whys and wherefores of my stay in prison have nothing to do with the subject at hand; but I must tell you how manuscripts go astray in the portfolios of the members of the Institute, although I cannot in truth conceive of such carelessness on the part of those who already have the death of Abel on their consciences. I do not want to compare myself with that illustrious mathematician but, suffice to say, I sent my memoir on the theory of equations to the Academy in February of 1830 (in a less complete form in 1829) and it has been impossible to find them or get them back. There are other anecdotes in this genre but I would be ungracious to recount them because, other than the loss of my manuscripts, those incidents do not concern me. Happy voyager, only my poor countenance saved me from the jaws of wolves. Perhaps I have already said too much for the reader to understand why, as much as I would have liked otherwise, it is absolutely impossible for me to embellish or disfigure this work with a dedication.[72]

The remainder of the *Préface* continues in much the same tone ("And thus it is knowingly that I expose myself to the laughter of fools"). Other of his writings are not dissimilar.[73] Among his papers is the picture of a bizarre, torsoless figure, captioned by Bourgne and Azra "Riquet à la Houppe."[74] The picture must have been drawn shortly before his death. It may be significant that Riquet à la Houppe was in French folklore a character, short, ugly, disdained by all, but nonetheless very clever.

VI. The Duel and Theories Surrounding It

We are almost at the end of this short story. Galois remained in Sainte-Pélagie without further recorded incident until March 16, 1832, when he was transferred to the pension Sieur Faultrier. Ironically enough, this was to prevent the prisoners from being exposed to the cholera epidemic then sweeping Paris. Galois was due to be given his freedom on April 29. From this point on, the historical record is very scanty. On May 25, Galois writes to his friend Chevalier and clearly alludes to a broken love affair:

My dear friend, there is a pleasure in being sad if one can hope for consolation; one is happy to suffer if one has friends. Your letter, full of apostolitic unction, has given me a little calm. But how can I remove the trace of such violent emotions that I have felt?

How can I console myself when in one month I have exhausted the greatest source of happiness a man can have, when I have exhausted it without happiness, without hope, when I am certain it is drained for life?[75]

The letter continues in similar tones. Galois goes on to say that he is disgusted with the world: "I am disenchanted with everything, even the love of glory. How can a world I detest soil me?"[76]

The next few days are a complete blank. On the morning of May 30, the famous duel took place. The previous evening, Galois wrote several well-known letters to his republican friends:

I beg patriots, my friends, not to reproach me for dying otherwise than for my country.

I die the victim of an infamous coquette and her two dupes. It is in a miserable piece of slander that I end my life.

Oh! Why die for something so little, so contemptible?

I call on heaven to witness that only under compulsion and force have I yielded to a provocation which I have tried to avert by every means. I repent in having told the hateful truth to those who could not listen to it with dispassion. But to the end I told the truth. I go to the grave with a conscience free from patriots' blood.

I would like to have given my life for the public good.

Forgive those who kill me for they are of good faith.[77]

Galois also writes another, similar letter to two republican friends, Napoleon Lebon and V. Delauney:

> My good friends,
>
> I have been provoked by two patriots . . . It is impossible for me to refuse.
>
> I beg your forgiveness for not having told you.
>
> But my adversaries have put me on my word of honor not to inform any patriot.
>
> Your task is simple: prove that I am fighting against my will, having exhausted all possible means of reconciliation; say whether I am capable of lying even in the most trivial matters.
>
> Please remember me since fate did not give me enough of a life to be remembered by my country.
>
> I die your friend.[78]

We will return to Galois's activities during this last night in due time. For now we discuss a few of the many theories which purport to explain the cause of this celebrated duel. There is perhaps enough in the two letters to raise suspicions of foul play. The attempts to make Galois the victim of royalists, a female agent provocateur, a prostitute, or a government conspiracy doubtlessly stem from these letters, for there is no other direct evidence in existence. Thus, we have the origin of Bell's assertion:

> What happened on May 29th is not definitely known. Extracts from two letters suggest what is usually accepted as the truth: Galois had run afoul of political enemies immediately after his release.[79]

The first statement is accurate, the second is not. Dupuy certainly believes the exact opposite, as will be seen shortly. Dupuy does mention that Alfred Galois, unjustifiably in his view, did maintain that his older brother was murdered. Because Bell "followed" Dupuy exclusively, one can only conclude that he took Alfred's position and termed it widely accepted or that he invented the whole thing.

Although Bell may have invented the theory, or merely propagated it to previously unattained heights, he is not its chief advocate. Infeld goes further. He assumes the "infamous coquette" was a female agent provocateur who set up Galois for the duel with a police agent. Infeld's evidence is by admission circumstantial. In addition to the bullet episode at Sainte-Pélagie it consists of the following:[80] the police were known to have used spies; the police broke up a meeting of the Society of Friends of the People the night before Galois's funeral; Police Chief M. Gisquet wrote in 1840 that Galois "had been killed by a friend"; police spies were unmasked in 1848, at which time a claim appeared in a journal that Galois "had been murdered in a so-called duel of honor"; Galois's brother Alfred always maintained that Evariste had been murdered; Galois was abandoned by his adversaries and his seconds and found by a peasant.

It should be noted that this evidence is consistent and does not contradict known facts. However, necessity does not follow from consistency. The bullet episode has already been discussed. It is true that the police used spies and that they were unmasked in 1848. We will return to this point below. Infeld does not mention that the newspapers announced Galois's funeral *before* the fact and explicitly named him as a member of both the Artillery of the National Guard and the Friends of the People. In any event, his membership in these organizations must have been widely known. One must weigh for oneself whether it is remarkable that police knew of republican meetings. Infeld finds it suspicious that the police chief, eight years after the fact, knew Galois had been "killed by a friend." He does not find it suspicious that Dumas knew more—precisely who that friend was. Dupuy feels that Alfred's position was the result of justifiable anger over his brother's death and points out some unlikely details Alfred attributed to the duel, such as stating that Evariste would have fired into the air. The assertion that Galois was abandoned to die, another of Alfred's claims, is also open to dispute. Dupuy mentions that one of the witnesses went to Galois's mother the following day to explain what had happened.[81] He feels, then, it was more likely that the witnesses were searching for a doctor when the peasant happened along. This explanation may be weak; nonetheless Infeld fails to mention that Mme. Galois was informed.

The remarks above are admittedly as circumstantial as the evidence. There is, however, more concrete evidence which weighs very heavily against the political conspiracy theorists: the identity of Pescheux d'Herbinville. More is known about him than his anonymity. He was, in fact, one of the nineteen republicans who were acquitted on charges of conspiring to overthrow the government in the trial spoken of earlier. Is there any reason to suspect the d'Herbinville was a police agent? The historian Louis Blanc, in his exhaustive *History of Ten Years*, writes:

> The trial gave rise to highly interesting scenes. In the sittings of the 7th of April, the president having reproached M. Pescheux d'Herbinville, one of the accused, with having had arms by him and with having distributed them, "Yes," replied the prisoner, "I have had arms, a great many arms, and I will tell you how I came by them." Then, relating the part he had taken in the three days, he told how, followed by his comrades, he had disarmed posts, and sustained glorious conflicts; and how, though not wealthy, he had equipped national guards at his own cost. There still burned in the hearts of the people some of the fire kindled by the revolution of July; such recitals as this fanned the embers. The young man himself, as he concluded his brief defense, wore a face radiant with enthusiasm and his eyes filled with tears.[82]

In addition, Blanc mentions the appearance of General Lafayette during the trial:

> The old general came to give his testimony in favor of the accused, almost all of whom he knew, and all saluted him from their places with looks and gestures of regard.[83]

D'Herbinville, it seems, was one of the heros of the hour. After the acquittal, the crowd pulled his coach through the streets of Paris "amid shouts of rapturous applause."

Bell, by not mentioning d'Herbinville at all, relieves himself of the difficulty of explaining why Galois should be killed in a political duel with a fellow republican or why d'Herbinville should be considered a political enemy. Infeld is in a more difficult position. Having acknowledged d'Herbinville's existence, he must explain why neither Dumas nor Blanc, both republicans, nor evidently the extremely liberal Lafayette* (assuming he knew d'Herbinville personally), nor, one would gather from Blanc's account, any republican in Paris, ever held any suspicions that d'Herbinville was an agent. Infeld talks at length about the 1848 unmasking of the police spies, but he does not mention the following extract from Dupuy:

> Pescheux was certainly not a "false-brother": all the men who acted as police agents during the reign of Louis-Phillipe were revealed in 1848 when Caussidière became chief of police, as witness Lucien de la Hodde.† If Pescheux were suspect, he would certainly not have been nominated as curator of the palace of Fontainbleu. It is absolutely necessary to discard the idea of police intervention and of a framed assassination.[84]

Thus there are some serious difficulties with the political enemies scenario. Infeld gets around this problem in characteristic fashion: in his bibliography he cites both Blanc and Dupuy as primary sources but *quotes neither*. In his Afterword, Infeld goes so far as to admit, "There is no reason to believe Pescheux d'Herbinville was a police agent." But then he goes on to say: "I believe there is enough circumstantial evidence to prove that the intervention of the secret police sealed Galois's fate. I do not believe it is possible to fit all the known facts without assuming Galois was murdered."[85]

It is left as an exercise for the reader to form a rebuttal to this statement. But in order to see just how far "known facts" can be stretched, we turn to Hoyle's version of the event. He writes:

> Such are the bare bones of the story of the life and death of Evariste Galois. The classical biography of Galois [he then references Dupuy], in an attempt to add flesh to these bones, suggests that he was done to death by royalist enemies, as does E. T. Bell in his book *Men of Mathematics*. There are dark hints that the release from prison was but a device for encompassing his death, a necessary preliminary to his being matched against a highly skilled assailant in royalist pay. But why should Galois feel it critical to his honor that he should accept the challenge of a right-wing agent, especially if the agent were a known marksman? Gallic logic suggests on account of a girl . . .[86]

*Lafayette had been considered republican enough to see his post of Commander of the National Guard dissolved after the events of December 1830.

† Hodde was a "republican" who was unmasked as a spy in 1848.

We first note the complete misrepresentation of Dupuy's position. If Hoyle is challenging Bell, and admittedly this is unclear, it seems to be on the extremely naive assumption that Galois would have known his opponent was a right-wing agent. Hoyle then goes on to dispose of the "infamous coquette" and propose his own theory:

> It is possible that the "infamous coquette" was the source of a purely personal quarrel, but it is the normal biological rule among mammals that sexual quarrels between two males cease as soon as one side seeks "accommodation." It is the normal rule that either party to such fights can simply walk away, which is just what Galois seems to have attempted to do.
>
> The more likely possibility is that Galois's habit of working mathematical problems in his head, his ability to think in parallel, caused serious animosities, and perhaps suspicions, to develop during the six months of imprisonment. There may have been suspicions that Galois was not wholly for the "cause," or even that he was an *agent provocateur* . . .[87]

Lincoln's remark comes to mind: "You can fool some of the people some of the time . . ." To suggest as Hoyle does that any republican in Paris suspected Galois after his expulsion from l'Ecole Normale, his Artillery activities, his threat to the King, his arrests, trials, sentencings, resentencings, and prison activities borders on the fantastical. This is in addition to the fact that two or three thousand republicans later attended the funeral of this supposed agent provocateur. One might equally well claim Lenin had been suspected of being a Menshevik.

As to Hoyle's bio-sociological theories, he is contradicted by the historical record. The greatest Russian poet, Alexander Pushkin, was killed in 1837 at the age of 37 in a duel over his wife. England's Lord Camelford was killed in a duel over a prostitute. As late as 1838 members of the American legislature were engaging in similar duels. Toward the end of the eighteenth century, during election season, approximately 23 duels *per day* were fought in Ireland *alone*, unlikely just for political reasons. In the last decades of the nineteenth century, Paris newspapers carried notices of the daily duels and their terms. These practices continued until World War I. The cause of such "affairs of honor" ranged from geese, to insults, to politics, to women.[88] Dupuy himself mentions that nothing was more common at the time in question than duels between republicans, and I think one may safely infer from his remarks that no one paid the slightest attention to them.[89]

However, argument by analogy is generally a weak policy when dealing with a specific case, and Hoyle's expansive pronouncements on the sexual behavior of mammals bring to mind further evidence with which anyone wishing to invoke a political cause for Galois's death must contend. This evidence consists primarily of two fragmentary letters written to Galois by one Mademoiselle Stéphanie D., who is none other than the "infamous coquette" over whom the duel was fought. Most authors have assumed her identity to be an absolute mystery and that she, like d'Herbinville, is an anonymous casualty of history. Dupuy apparently was unaware of the letters or chose not to publish them. Bell and Hoyle never mention her name. Infeld calls her Eve Sorel (perhaps inspired by Stendhal). This is a strange state of affairs, for the letters were published in Tannery's 1908 edition of Galois's papers. Tannery does not affix a name to the author of these fragments; it is left for the 1962 edition of Bourgne and Azra to attempt an identification. One can understand why Bell and Infeld did not mention her name since Tannery did not provide it. Hoyle does not have such an excuse, his book being published in 1977. One cannot understand why these letters are never mentioned by anyone, especially by Bell and Infeld who cite Tannery as a major source for Galois's manuscripts.

The letters, as they exist, are copies made by Galois himself on the back of one of his papers.[90] The copies contain gaps, which may indicate he had previously torn up the originals and could not completely reconstruct them. More likely, Galois purposely omitted any incriminating or personally distasteful segments. I say this because some words in the French versions are broken in half; one generally does not remember only half a word. Galois has certainly obliterated Stéphanie's last name in a fit of anger. Due to the fragmentary nature of these letters their translation has proved difficult and may be uncertain in places. Where impossible to translate we have allowed the original French to stand. Letter I:

> Please let us break up this affair. I do not have the wit to follow a correspondence of this nature but I will try to have enough to converse with you as I did before anything happened. Here is Mr. the *en a* *qui* *doit vous* *qu'a* me and do not think about those things which did not exist and which never would have existed.
>
> Mademoiselle Stéphanie D
> 14 May 183—

Letter II:

> I have followed your advice and I have thought over what has happened on whichever denomination it may have happened between us. In any case, Sir, be assured there never would have been more. You're assuming wrongly and your regrets have no foundation. True friendship exists nearly only between people of the same sex particularly of friends. full in the *vacuum* that the absence of all feeling of this kind . . . my trust . . . but it has been very wounded . . . you have seen me sad you have asked the reason; I answered you that I had sorrows that one had inflicted upon me. I had thought that you would take this as anyone in front of whom one drops a word for these one is not The calm of my thoughts leaves me to judge the persons that I usually see without much reflection; this is the reason that I rarely regret having been wrong in my judgment of a person. I am not of your opinion *les sen* *plus que* *les a* *exiger* *ni se* thank you sincerely for all those who you would bring down in my favor.

These are highly tantalizing morsels, but is there anything else known about the author? Indeed there is. C. A. Infantozzi has examined the original of the first letter.[91] With the help of a magnifying glass he was able to discern Stéphanie's full signature under Galois's erasures: Stéphanie Dumotel. Further archival investigation by Infantozzi shows she was Stéphanie-Félicie Poterin du Motel, daughter of Jean-Louis Auguste Poterin du Motel, a resident physician at the Sieur Faultrier, where Galois stayed the last months of his life. In 1840 Stéphanie married Oscar-Théodore Barrieu, a language professor. Any presumption that she was a prostitute must at this point be discarded as a complete figment of Bell's imagination.

The establishment of Stéphanie's identity unfortunately does not conclusively establish what in actuality did occur. From Stéphanie's second letter it is not difficult to infer that Galois took some song of sorrows on her part too seriously and himself provoked the duel. On face value she certainly seems an unwilling participant in whatever transpired. On the other hand, we have a curious passage from Dupuy, once again not quoted by the other authors. During the course of his researches, Dupuy had asked Galois's cousin if he knew the cause of the duel.

> His cousin, M. Gabriel Demante, writes me that at a last meeting [with Stéphanie?] Galois found himself in the presence of a supposed uncle and a supposed fiancé, each of whom provoked the duel.[92]

It is difficult to say in which direction this passage points, but in weighing its importance, one should keep in mind Dupuy's own skepticism of anything the Galois family said concerning the duel.

With this passage, all the evidence pertaining to the duel which I have found to date has been presented. One can read the circumstantial evidence as Infeld does to arrive at a conspiracy. No known facts conclusively refute this interpretation. But it must be reemphasized that there is absolutely no direct evidence that such is the case. Furthermore, there is the testimony of several men, two of whom were republicans, that d'Herbinville was *not* a police agent. In addition, there is the identity of Stéphanie, who was simply the daughter of a physician who happened to live and work at the pension where Galois was staying. To suggest she was an agent provocateur somehow planted there to entrap Galois becomes a baroque, if not byzantine, invention.

If one chooses to reject the conspiracy theory, a fairly consistent picture of a personal quarrel emerges. We have Galois's unhappy letter to Chevalier of May 25. His famous cry, "I am the victim of an infamous coquette and her two dupes," may mean exactly what it says, with suitable allowances for Galois's usual withering tone. The excerpt from Dupuy quoted above is certainly consistent with "two dupes." Galois himself writes, "Forgive those who kill me for they are of good faith," i.e., they are not political enemies. And we must remember, in his own eyes, Galois

was exceedingly honest. If he felt any treachery were involved, we can be sure he would have said as much. The two letters from Stéphanie are perhaps the strongest argument for a personal quarrel. Those dissenting could of course take the extreme position that she was a very good actress.

In addition, there is the more difficult question of psychology. Galois's writings are at times unquestionably violent, and equally violent erasures are preserved on his manuscripts. He was arrested twice for dangerous actions which might have easily been avoided by a more prudent individual, or perhaps in a more prudent age. Raspail writes that Galois attempted suicide in prison, and of course there is Galois's own prophecy, not inconceivably self-fulfilling, that he would be killed in an affair of honor. It is not terribly difficult to believe that such a troubled young man in such a turbulent time could have ended his life in a duel.

In this scenario, the role of Pescheux d'Herbinville admittedly remains unclear. Was he a "supposed" fiancé or a real fiancé whom Galois's cousin took for supposed? Did Galois's cousin invent this epithet, or was d'Herbinville simply involved in a stupid quarrel? I have no answers to these questions.

The point I wish to make now for the interested historian is that, although in 1982 Galois will have been dead 150 years, the investigation of his death has been closed prematurely. D'Herbinville should be traced to see if any letters exist which might shed some light on the matter.* If he was in prison with Galois, a background for the quarrel might be established. Letters of Stéphanie or her husband might be extant and could conceivably mention the duel. Dupuy remarks cryptically that Raspail as well as all the republicans knew the cause of the duel. Raspail became a famous politican. Perhaps there is a clue in his correspondence.

These avenues are still open for those who are interested. They have been neglected only because of the intentional or unintentional omission of information by those who have previously written on Galois. We will return to the question of scholarship after disposing of the remaining myths concerning the night before the duel.

VI. The Last Night

We saw in the introduction how Bell all but states outright that Galois committed his theory of equations to paper the night before he was shot. James R. Newman repeats this as an assertion, and the vision of the doomed boy, sitting by candlelight, feverishly bringing group theory into the world seems to be the major myth which most scientists harbor concerning Galois. This is again due to Bell's embellishment of Dupuy, who in this instance is sufficiently romantic of his own accord. But as has already been detailed at great length, Galois had been submitting papers on the subject since the age of 17. The term "group," used in the sense of "group of permutations" is used in all of them. During the night before the duel, in addition to the letters already quoted, Galois wrote a long letter to his friend Chevalier.[93] He begins:

> My Dear Friend,
>
> I have made some new discoveries in analysis.
>
> The first concern the theory of equations, the others integral functions.
>
> In the theory of equations I have researched the conditions for the solvability of equations by radicals; this has given me the occasion to deepen this theory and describe all the transformations possible on an equation even though it is not solvable by radicals.
>
> All this will be found here in three memoirs.

Galois then goes on to describe and elucidate the contents of the memoir which was rejected by Poisson, as well as subsequent work. Galois had indeed created a field which would keep mathematicians busy for hundreds of years, but not "in those last desperate hours before the dawn." During the course of the night he annotated and made corrections on some of his papers. He comes across a note that Poisson had left in the margin of his rejected memoir:[94]

*A perusal of the standard biographical encyclopedias has failed to reveal any further information on d'Herbinville.

The proof of this lemma is not sufficient. But it is true according to Lagrange's paper, No. 100, Berlin 1775.

Galois writes directly beneath it:

This proof is a textual transcription of that which we gave for this lemma in a memoir presented in 1830. We leave as an historic document the above note which M. Poisson felt obliged to insert. (Author's note.)

A few pages later,[95] Galois scrawls next to a theorem:

There are a few things left to be completed in this proof. I have not the time. (Author's note.)

Galois penned this famous inscription only once during the course of the night. It is unfortunate he tarnished some of the romance by including his parenthetical "Author's note." Galois ends his letter to Chevalier with the following request:

In my life I have often dared to advance propositions about which I was not sure. But all I have written down here has been clear in my head for over a year, and it would not be in my interest to leave myself open to the suspicion that I announce theorems of which I do not have complete proof.

Make a public request of Jacobi or Gauss to give their opinions not as to the truth but as to the importance of these theorems.

After that, I hope some men will find it profitable to sort out this mess.

I embrace you with effusion. E. Galois.[96]

And that was the end. The funeral was to be held on June 2. During the previous evening, the police broke up a meeting of the Society of Friends of the People on the pretext that the republicans were planning a demonstration for Galois's funeral. Thirty of those present were arrested. The next day, two or three thousand republicans were present at the services. Galois's body was interred in a common burial ground of which no trace remains today.

Later, Evariste's brother Alfred and his devoted friend Chevalier would laboriously recopy the mathematical papers and submit them to Gauss, Jacobi, and others. By 1843 the manuscripts had found their way to Liouville, who, after spending several months in the attempt to understand them, became convinced of their importance. He published the papers in 1846.

There exist many fragments which indicate Galois carried on his mathematical researches, not only while in prison, but right up until the time of his death. The fact that he could work through such a turbulent life is testimony to the extraordinary fertility of his imagination. There is no question that Galois was a great mathematician who developed one of the most original ideas in the history of mathematics. The invention of legends does not make him any greater.

VII. Harsher Words

The account of Galois's life given here has not been entirely complete. There are more documents, letters, and events. No doubt I will shortly be exposed for having selectively presented evidence. The purpose of this paper, however, has not been one of completeness, nor entirely one of biography. No, the purpose has been to show that something is wrong. Two highly respected physicists and an equally well-known mathematician have invented history.

Bell's account, by far the most famous, is also the most fictitious. It is a myth devoid of such complications as a protagonist who is faulted as well as gifted. It is myth based on the stereotype of the misunderstood genius whom the conservative hierarchy is out to conquer. As if the befuddled hierarchy is generally organized well enough for persecution. It is a myth based on a misunderstanding of the method by which a scientist works: as if a great theory could be written down coherently in a single night.

It is unclear how far one can go in forgiving Bell. Surely all his mistakes could not result from a poor knowledge of French. No, I believe consciously or unconsciously Bell saw his opportunity to create a legend. The details which are absent in his account, such as Dumas at the banquet, such as d'Herbinville, such as the suicide attempt and Raspail, are those details which lend a concreteness and a humanness to Galois's life which a legend must not have. Unfortunately, if this was Bell's intent, he succeeded. After hearing of my investigation, physicists and mathematicians

all open conversations with me with the same question: "Did Galois really invent group theory the night before he was killed?" No, he didn't.

Infeld presents far more details. He is not interested in making Galois a legend. He does intend to make Galois a hero of the people. Politics is the guiding principle for Infeld. His book might be termed the proletarian interpretation of Galois; certainly parts of it read like the local Workers' Party publication. Infeld is very good at covering his tracks. To delete a phrase here, a paragraph there, a counterargument in between, is all that is necessary to create conspiracy from chaos.

As to Hoyle's motives, we can only take him at his word: He describes at length how as a child he was taught arithmetic by his mother, how he became proficient at mathematics, and how school for him became an excruciating bore. Hoyle was forced to learn to "think in parallel" in order to fool the teacher into believing he paid attention in class. He then writes, "I mention these personal details because I believe they cast some light on the mysterious death of the French mathematician Evariste Galois." Further comment seems unnecessary.

Dupuy seems to have less of a vested interest. I assume he included all the documents known to him at the time. If not, then he too should be scrutinized more carefully. He does seem *a priori* unwilling to accept conspiracy theory.

At the very least, the three twentieth-century authors are guilty of distorting Dupuy's account and even falsifying it. In each case the story of Galois has been used to put a stamp of approval on the author's personal theories. Indeed, all history is interpretive. But if we do not approve, we understand the liberty: Galois, like Einstein, has passed into the public domain. No act or anecdote attributed to him is too outrageous to be given consideration. There is a closer analogy from farther afield. The Russian composer Reinhold Glière once wrote a symphony, his third, which ran well over an hour. Stokowski—the story goes—worked with Glière to edit the score down to manageable length. Since then, every conductor presents his own edition. I do not know if I have ever heard the original.

The investigations of Galois discussed here have told us less about the man than about his biographers. The misfortune is that the biographers have been scientists. Because they appreciate his genius a century after its undisputed establishment, anyone who did not recognize it at the time is condemned. "In all the history of science," writes Bell, "there is no completer example of the triumph of crass stupidity over untamable genius." "Is it possible to avoid the obvious conclusion," asks Infeld, "that the regime of Louis-Phillipe was responsible for the early death of one of the greatest scientists who ever lived?" The underlying assumption is apparent: Galois was persecuted because he was a genius, and all scientists, to a greater or lesser degree, understand that genius is not tolerated by mediocrity. From this point of view, a genius must be recognized as such even when standing drunk on a banquet table with a dagger in his hand. Anyone who does not recognize him becomes a fool, an assassin, or a prostitute. This is a presumption of the highest arrogance. Scientists should not be so enamored of themselves.

Acknowledgements

I would like to thank Leonard Gillman for suggesting that I consolidate my researches on Galois into this article. Thanks to Cecile DeWitt-Morette for the gracious gift of her time in translating, and most of all to Marc Henneaux for his translations, patience, discussions and enthusiasm.

Notes

1. Freeman Dyson, *Disturbing the Universe* (New York: Harper and Row, 1979), p. 14.
2. E. T. Bell, *Men of Mathematics* (New York: Simon and Schuster, 1937), p. 375.
3. James R. Newman, *The World of Mathematics* (New York: Simon and Schuster, 1956), vol. 3, p. 1534.
4. *Checklist of the Bullitt Collection of Mathematics* (University of Louisville, 1979).
5. Leopold Infeld, *Whom the Gods Love: The Story of Evariste Galois* (New York: Whittlesey House, 1948).
6. Fred Hoyle, *Ten Faces of the Universe* (San Francisco: W. H. Freeman, 1977), Chap. 1.

7. Paul Dupuy, La Vie d'Evariste Galois, *Annales de l'Ecole Normale*, 13 (1896) 197–266.
8. Bell, p. vii.
9. Jules Tannery, ed., *Manuscrits d'Evariste Galois* (Paris: Gauthier-Villars, 1908).
10. Robert Bourgne and J. P. Azra, eds., *Ecrits et Mémoires Mathématiques d'Evariste Galois*: *Edition Critique Intégrale de ses Manuscrits et Publications* (Paris: Gauthier-Villars, 1962).
11. Alexandre Dumas, *Mes Mémoires* (Paris: Editions Gallimard, 1967), vol. 4, Chap. 204.
12. François Vincent Raspail, *Lettres sur les Prisons de Paris* (Paris, 1839), vol. 2.
13. Bell, p. 362.
14. Bell, p. 363.
15. Dupuy, p. 203.
16. Bell, p. 364. Compare with Dupuy, p. 205.
17. Dupuy, pp. 255–256.
18. Dupuy, pp. 254–255.
19. Dupuy, p. 256.
20. Dupuy, p. 209; Bell, p. 368.
21. Infeld, p. 306.
22. René Taton, "Sur les relations scientifiques d'Augustin Cauchy et d'Evariste Galois," *Revue d'Histoire des Sciences*, 24 (1971) 123.
23. Bell, p. 368.
24. Bell, p. 369.
25. Dupuy, p. 211.
26. Bell, p. 371.
27. Dupuy, p. 217.
28. Bourgne and Azra, p. 27. See also pp. 21–25.
29. I assume here, as elsewhere, the chronology given by Bourgne and Azra, pp. xxvii–xxxi.
30. Bell, p. 370.
31. Taton, p. 134.
32. Taton, p. 139.
33. Bourgne and Azra, p. xxviii.
34. Taton, p. 138.
35. See, for example, Lilian Lieber, *Galois and the Theory of Groups* (1932).
36. Bourgne and Azra, p. xxviii.
37. Bell, p. 366.
38. Dupuy, p. 221, and Bourgne and Azra, p. xxix.
39. Bourgne and Azra, p. 462. Also Dupuy, p. 225. Translated in part by Infeld, p. 155.
40. Dupuy, pp. 227–228.
41. This account is based on information from Louis Blanc, *History of Ten Years* (London: Chapman and Hall, 1844). A consistent account, also based on Blanc, is given by Infeld, Chap. 5.
42. Dupuy, p. 234.
43. Bell, p. 372.
44. C. Henry, "Manuscrits de Sophie Germain," *Revue Philosophique*, 8 (1879) 631.
45. See, for example, his Afterword.
46. Infeld, p. 169.
47. Dumas, p. 331.
48. Bell, p. 372.
49. Dumas, pp. 332–333.
50. Dupuy, pp. 234–235.
51. Dupuy, p. 235.
52. Dupuy, p. 234.
53. Dupuy, p. 247.
54. Because of a change of libraries, this account of the trial is based on a different edition of Dumas's memoirs: Alexandre Dumas, *Mes Mémoires*, (Paris: Union Générale d'Editions), vol. 2, chap. 37. (No copyright date given.)
55. Infeld, p. 230.
56. See, for example, T. E. B. Howarth, *Citizen King*: *The Life of Louis-Phillipe* (London: Eyre & Spottiswoode, 1961).
57. Dupuy, p. 238.
58. Bell, p. 378.
59. Dupuy, p. 238.
60. Dora B. Weiner, *Raspail*: *Scientist and Reformer* (New York: Columbia University Press, 1968).

61. Raspail, p. 84.
62. Raspail, p. 89.
63. Raspail, p. 90.
64. Bell, p. 374.
65. Dupuy, p. 245.
66. Bell, p. 374.
67. Raspail, pp. 117–118.
68. Raspail, p. 118. Also discussed in Dupuy, p. 243.
69. Infeld, p. 308.
70. From Joseph Bertrand, "La vie d'Evariste Galois par P. Dupuy," *Eloges Académiques*, Paris, 1902, pp. 329–345. Here, I have quoted the letter as published by Infeld, p. 230.
71. Bell, p. 371.
72. Bourgne and Azra, pp. 3–11.
73. See again Bourgne and Azra, pp. 21–27.
74. Bourgne and Azra, facsimiles.
75. Bourgne and Azra, pp. 468–469.
76. Bourgne and Azra, p. 469.
77. Bourgne and Azra, p. 470.
78. Bourgne and Azra, p. 471.
79. Bell, p. 375.
80. Infeld, pp. 308–311.
81. Dupuy, pp. 247–248.
82. Blanc, p. 431.
83. Blanc, p. 431.
84. Dupuy, p. 247.
85. Infeld, p. 310.
86. Hoyle, p. 14.
87. Hoyle, p.15.
88. See, for example, Charles Mackay, *Extraordinary Popular Delusions and the Madness of Crowds* (USA: Noonday Press, 1932), Chap. "Duels and Ordeals"; and Roger Shattuck, *The Banquet Years* (New York, Vintage Books, 1968), Chap. 1.
89. Dupuy, p. 247.
90. The letters and the description of them are in Bourgne and Azra, pp. 489–491.
91. C. A. Infantozzi, "Sur la mort d'Evariste Galois," *Revue d'Histoire des Sciences*, 21 (1968) 157.
92. Dupuy, p. 246.
93. Bourgne and Azra, p. 173.
94. Bourgne and Azra, p. 48 and facsimiles.
95. Bourgne and Azra, p. 54 and facsimiles.
96. Bourgne and Azra, p. 185.

MISCELLANEA

67. *Simplicio.* I am really beginning to understand that logic, although a most excellent instrument for organizing our reasoning, cannot match the acuteness of geometry in directing the mind toward discovery.

Sagredo. It seems to me that logic can tell us whether arguments and proofs are conclusive; but that it can tell us how to discover conclusive arguments and proofs, that I really do not believe.

—Galileo Galilei, *Discorsi e dimostrazioni matematiche intorno à due nuove scienze* (1638). (*Le Opere di Galileo Galilei*, Edizione Nazionale, Firenze, 1898, vol. 8, p. 175.)

These photographs of four mathematicians who are famous now were taken about 1938 and 1939. Two of the subjects are still alive. They represent, but not necessarily in order, analysis, number theory, statistics, and topology. The secret is revealed on p. 133.

Let us define for each region of this lattice a *shell index*. The original triangle has a shell index of 0. The three hexagons bordering it have shell index 1. The twelve triangles bordering these hexagons have shell index 2, etc. The nth shell consists of

(1) $3n$ hexagons, if n is odd;

(2) $6n$ triangles, if n is even.

If P_0 is any point in shell n, then its orbit intersects each region in the nth shell and no others. If n is even, this orbit has exactly one point in each region; hence it has $6n$ points, i.e., $P_{6n} = P_0$. Thus for the stippled triangles (shell 2), $P_{12} = P_0$; for the hatched triangles (shell 4), $P_{24} = P_0$; etc.

If n is odd, there is a special case: if P_0 is the center point of a hexagon (each of which is symmetric under a half-turn about its center), then its orbit has again exactly one point in each region of the shell—namely, the midpoints—so $P_{3n} = P_0$. Thus for the points denoted $\triangle$ in the figure (shell 1), $P_3 = P_0$; for those denoted x (shell 3), $P_9 = P_0$; for those denoted $\circ$ (shell 5), $P_{15} = P_0$; etc.

If P_0 is any other point in a hexagon, then its orbit includes two symmetrically located points in each hexagon. For such points, $P_{6n} = P_0$. Thus for points in the innermost three hexagons (excluding the $\triangle$'s), $P_6 = P_0$. For points in the next ring of nine hexagons (excluding the $\times$'s), $P_{18} = P_0$. For points in the 15 hexagons in the outermost ring shown (excluding the $\circ$'s), $P_{30} = P_0$.

The region in which P_n lies and the region next to the region in which P_{n+1} lies are always opposite (with respect to Δ). Thus the sequence seesaws around the shell. The P_n with odd indices make a continuous circuit around the shell, moving always to contiguous regions, and so do those with even indices. Typical sequences are illustrated in shells 2 and 3.

Rank 1 Matrices $A, B, A + B$

E 2851 [1981, 672]. *Proposed by Peter Ungar, New York University.*

Suppose all three matrices, A, B, $A + B$ have rank 1. Prove that either all the rows of A and B are multiples of one and the same row vector v or else all the columns of A and B are multiples of one and the same column vector w.

Solution by Noel Glick, student, Brooklyn College, New York. Since A has rank 1, clearly all the rows are multiples of one row vector $v \neq 0$. A similar remark applies to B with corresponding row vector w. Thus, if $A_i(B_i)$ denotes the ith row of $A(B)$ we have $A_1 = c_1v$, $A_2 = c_2v$, $A_3 = c_3v, \ldots$, and $B_1 = d_1w$, $B_2 = d_2w$, $B_3 = d_3w, \ldots$ (c_j and d_j are constants). Assume that w is not a multiple of v. It follows that the column vector $c = (c_1, c_2, c_3, \ldots)$ is a multiple of the column vector $d = (d_1, d_2, d_3, \ldots)$, otherwise $A + B$ would have at least 2 independent rows. But this means that the columns of A, B, and $A + B$ are all multiples of the column vector c.

Duarte mentions the generalization (Marcus, *Finite-dimensional Multilinear Algebra*, Dekker, 1973): Any nonzero $m \times n$ matrix A over a field can be written as a sum of r rank one matrices:

$$A = \sum_{t=1}^{r} [c_{ti} d_{ti}].$$

Moreover, A has rank r if and only if each of the two sets of vectors $\{(c_{t1}, \ldots, c_{tm})\}$, $\{(d_{t1}, \ldots, d_{tn})^*\}$, $t = 1, \ldots, r$ is a linearly independent set.

Also solved by A. L. Duarte (Portugal), M. Fields (student), G. Gagola, F. Gerrish (United Kingdom), V. Hernandez (Spain), J. T. Holmes, M. Josephy (Costa Rica), M. F. Kruelle (student), J. R. Kuttler, A. Lavin (student), O. P. Lossers, Jr. (Netherlands), M. Marcus, N. Passell, J. F. Queiró (Portugual), S. Ricci, G. S. Rogers, H. Schwerdtfeger (Canada), J. Suck (Germany), W. V. Webb, H. Wolkowicz (Canada), Y.-L. Wong (Hong Kong), P.-Y. Wu (China).

ANSWERS TO "PHOTOS" ON PAGE 107

Top left: A. S. Besicovitch; top right: D. H. Lehmer; bottom left: N. E. Steenrod; bottom right: J. W. Tukey.

APPROXIMATIONS TO THE BANZHAF INDEX OF VOTING POWER

SAMUEL MERRILL III
Department of Mathematics and Computer Science, Wilkes College, Wilkes-Barre, PA 18766

1. Introduction. Bernard Grofman devotes a recent MONTHLY article [**4**] to substantiating an assertion by Justice Harlan concerning the sensitivity of the absolute Banzhaf index of voting power to minor variations in assumptions. Although Grofman's calculations are correct, they pertain only to the *absolute*, not the *relative*, Banzhaf index and only to unweighted voting. It is the relative index that has been used to shed light on the relative power of voters in relation to the one man, one vote principle (see e.g., Banzhaf [**1**], Brams [**2**], Lucas [**5**], and Straffin [**9**]). Furthermore, since under unweighted voting, all voters possess the same relative power, only weighted voting leads to nontrivial results about relative power.

Recall that for a political body with n voters, the absolute Banzhaf index (see Banzhaf [**1**]) of a focal voter is the probability that the outcome be reversed if the focal voter should reverse his vote, under the assumption that voters vote independently for one of two alternatives A and B, each with probability $p = .5$. The relative Banzhaf index for the focal voter is obtained by normalizing so that the sum (over the n voters) of the indices is unity.

In his dissenting opinion in *Whitcomb* v. *Chavis*, Justice Harlan asserted that for $n = 300{,}000$, if the probability p were changed from .5 to .505, then each individual's voting power would be reduced by a factor of approximately 1,000,000. Although this statement is valid for the absolute index, the relative index for the case of n identical voters is the same for all voters and hence completely insensitive to alteration of the value of p. However, for weighted voting, relative Banzhaf indices of two voters may differ vastly from the ratio between their weights even for $p = .5$.

2. An Approximation for Unweighted Voting. Before proving a result for weighted voting, I will give a simple approximation for the absolute Banzhaf index for each of $n + 1$ identical voters when p is near .5. The technique will then be extended to weighted voting. Note that the number of votes cast for alternative A before counting the vote of the focal voter is a binomial random variable X with parameters n and p. By the central limit theorem, if $n \geqslant$ about 25, X may be approximated by a normally distributed random variable with parameters $\mu = np$ and $\sigma^2 = np(1 - p)$. Set $p = .5 + \delta$ and let f_δ denote the density of the approximating normal. The probability that the focal voter will be decisive in the sense described above is

$$P[X = n/2] \cong \int_{n/2-1/2}^{n/2+1/2} f_\delta \cong f_\delta(n/2). \tag{1}$$

The factor by which the absolute Banzhaf index is reduced when $p = .5$ is changed to $p = .5 + \delta$ is thus

$$R_\delta = f_0(n/2)/f_\delta(n/2) \cong \exp(2\delta^2 n) \tag{2}$$

for small δ.

For Harlan's example ($n = 300{,}000$, $\delta = .005$), $R_\delta = \exp(15) \cong 3.27 \times 10^6$. Harlan also states that if 15,000 of the voters are committed to A and 10,000 to B, then power is reduced by a factor of 1.2×10^{20}. This scenario is equivalent to setting $n = 275{,}000$ and $p = 27/55$ or $\delta = -1/110$. By (2), $R_\delta = \exp(45.\overline{45}) \cong 5.5 \times 10^{19}$, which is close to Harlan's value.

3. Approximations for Weighted Voting. Under weighted voting for $n + 1$ voters, the ith voter's vote is given a weight of w_i. I will consider in detail only the case in which all $w_i = 1$ except for w_1, which is greater than 1. This case illustrates some of the difficulties that may arise. Let $p = .5 + \delta$ and note that the sum of the weights is $n + w_1$. The probability P_1 that the first voter be decisive can be shown to be given by the approximation

$$P_1 \cong \int_{(n-w_1)/2}^{(n+w_1)/2} f_\delta, \tag{3}$$

by an argument similar to that used to derive (1) (see Owen [**7**] or Merrill [**6**]). Here f_δ is the approximating normal for the number of minor voters selecting alternative A. Now let P_i denote the probability of being decisive for any one of the small voters. The total weighted vote for alternative A can be approximated by a mixture of two normals each associated with one of the two possible votes by voter 1. Accordingly,

$$P_i \cong p f_\delta((n - w_1)/2) + (1 - p) f_\delta((n + w_1)/2). \tag{4}$$

Note that P_1 and P_i are absolute Banzhaf indices; whereas the ratio P_1/P_i measures the relative Banzhaf power between voter 1 and any one of the small voters. The following theorem provides information concerning this ratio.

THEOREM 1. *Suppose there are $n + 1$ voters, with voter 1 having weight w_1 and each of the other voters having weight 1. Assume that $\delta = 0$ (i.e., $p = .5$).*
(i) *If $w_1 \leqslant \sqrt{n}$, then $P_1/P_i \leqslant \sqrt{e}\, w_1$.*
(ii) *If $w_1 = k\sqrt{n}$ and $k \geqslant 1$, then $P_1/P_i \geqslant (w_1/2k)\exp(k^2/2)$.*

Proof. Suppose first that $w_1 \leqslant \sqrt{n}$. By (3),

$$P_1 \leqslant w_1 f_0(n/2) \cong w_1\sqrt{2/n\pi}, \tag{5}$$

since f_0 attains its maximum at $n/2$. Since f_0 is symmetric about $n/2$, and has parameters $\mu = n/2$ and $\sigma^2 = n/4$, (4) implies

$$P_i \cong f_0((n + w_1)/2) \cong \sqrt{2/n\pi}\exp\left(-\tfrac{1}{2}\left(w_1/\sqrt{n}\right)^2\right) \geqslant \sqrt{2/n\pi}\exp\left(-\tfrac{1}{2}\right). \tag{6}$$

Relation (i) follows from (5) and (6).

Now assume $w_1 = k\sqrt{n}$. If $k \geqslant 1$, $w_1/2 \geqslant \sqrt{n}/2 = \sigma$. Hence by (3), $P_1 \geqslant \Phi(1) - \Phi(-1) > \frac{1}{2}$, where Φ denotes the standard normal distribution function. Arguing as before,

$$P_i \cong \sqrt{2/n\pi}\exp(-k^2/2) = (k/w_1)\sqrt{2/\pi}\exp(-k^2/2), \tag{7}$$

so that relation (ii) follows by direct calculation.

Thus the relative Banzhaf power between voter 1 and any one of the small voters remains relatively close to w_1 as long as w_1 does not exceed $\sqrt{n}$; as w_1 grows larger, relative power increases exponentially. For example, if $n = 300{,}000$ and $w_1 = 3000$ (so that $k^2 = 30$), then P_1/P_i exceeds w_1 by a factor of at least 3×10^5. Note that Banzhaf's original assumption that $p = .5$ did not need to be altered to obtain this result.

If instead there are *two* major voters with identical weights w, similar arguments show that the ratio between power and weight is approximately the same for major and minor voters as long as $w \leqslant \sqrt{n}$. However, as w grows larger, the power to weight ratio approaches zero for the two major voters but not for the minor voters. This is the simplest example of the curious pitfall points determined by Shapley and Dubey [**8**] in which the major voters are all destroyed in the limit. Shapley and Dubey provide an exhaustive treatment of the asymptotic properties of the Banzhaf index.

4. Modifications of the Banzhaf Index. In closing I would like to mention two modifications of the Banzhaf measure of power which yield more realistic estimates of the absolute power of voters for different values of p. In [**6**] I argue empirically that a large electorate, such as that of a U.S. state, can be approximated by a moderate number, say K, of independently voting blocs. The value of K should be chosen so that $p(1 - p)/K$ is in accord with the sample variance of the proportion of the vote received by one party over time. Empirical evidence [**3**], [**6**] suggests that K is of the order of 10's or 100's for most U.S. states. If, e.g., $K = 100$ and I replace (2) with

$$R_\delta = \exp(2\delta^2 K), \tag{2'}$$

then $R_{.01} = \exp(.02) = 1.02$, $R_{.1} = \exp(2) = 7.4$, and $R_{.2} = \exp(8) \cong 3000$. Thus in practice a large perturbation of p is required to produce a major reduction in absolute voting power in an unweighted model.

Straffin [9] investigates the limited independence of voters in real situations and presents a partial homogeneity model which yields power indices which do not require statistical data. His model incorporates the Shapley model within blocs and the Banzhaf model among blocs. I believe that Straffin's model can play an important role in assessing voting power when the strict independence assumption of Banzhaf's original model is not tenable.

References

1. J. Banzhaf, One man, 3.312 votes: A mathematical analysis of the Electoral College, Villanova Law Rev., 13 (1968) 304–332.
2. S. J. Brams, Game Theory and Politics, Free Press, New York, 1975.
3. C. Colantani, T. Levesque, and P. Ordeshook, Rejoinder to "Comment" by S. J. Brams and M. D. Davis, Amer. Political Sci. Rev., 69 (1975) 157–161.
4. B. Grofman, Fair apportionment and the Banzhaf index, this MONTHLY, 88 (1981) 1–5.
5. W. F. Lucas, Measuring Power in Weighted Voting Systems, Tech. Rep. no. 227, Department of Operations Research, Cornell University, Ithaca, 1974.
6. S. Merrill, Citizen voting power under the Electoral College: A stochastic model based on state voting patterns, SIAM J. Appl. Math., 34 (1978) 376–390.
7. G. Owen, Multilinear extensions and the Banzhaf value, Naval Res. Logist. Quart., 22 (1975) 741–750.
8. L. S. Shapley and P. Dubey, Mathematical Properties of the Banzhaf Power Index, Math. of Operations Research, 4 (1979) 99–131.
9. P. Straffin, Homogeneity, independence, and power indices, Public Choice, 30 (1977) 107–118.

LONG PROOFS

F. H. NORWOOD
Department of Mathematics, Lehigh University, Bethlehem, PA 18015

In every sufficiently complicated mathematical system, say a system which includes Peano arithmetic, there are infinitely many theorems. To confirm our intuition that the number of theorems is infinite, we can simply consider theorems of the form "The integer p is prime," in which the symbol p is replaced by any prime number. Now a theorem is at least implicit in its proof, and we can make this explicit by requiring that a proof begin with a statement of the theorem, so that distinct theorems have distinct proofs, and there are therefore infinitely many distinct proofs in our system.

This could be a starting point for a discussion of the Gödel incompleteness theorem [2], but there is a simpler theorem to be obtained from the same beginning.

THEOREM. *There are true theorems with long proofs. That is to say, for any natural number N there is a true theorem whose shortest proof contains more than N alphanumeric characters.*

Proof. For any N there are only finitely many strings of N or fewer alphanumeric characters.

Frederick Norwood got his Ph.D. from the University of Southwestern Louisiana, working under Professor Wilbur Whitten. This paper was written while he was a member of the Institute for Advanced Study. His research interests include the theory of knots and the classification of three-manifolds. Under the name Rick Norwood he writes science fiction.—*Editors*

Therefore, there are only finitely many theorems which have any proof of length N or less. Therefore there are infinitely many theorems whose shortest proof has length greater than N.

A few comments may be in order.

1. I restrict myself to alphanumeric characters to avoid the following case. Imagine a mathematical system which allows infinitely many distinct characters (as distinct from ordinary systems, which only allow infinitely many distinct strings of characters). For example, imagine a system in which a slash (/) had a distinct meaning which depended on its length. We have, then, uncountably many distinct symbols in this system, and one can conceive of a system in which there are infinitely many theorems, each with a proof one symbol long, the proof consisting of a single symbol which stands, uniquely, for the proof of that theorem.

The human mind can imagine such a system, but it cannot comprehend such a system, since from these infinitely many characters the human eye can only distinguish finitely many equivalence classes of characters at best, even if we could avoid ambiguity.

Therefore it seems reasonable to limit ourselves to finitely many symbols and, having done so, the alphanumeric symbols are handy. We could make do with only 0 and 1, but there is no need. Nor is there any logical problem in writing out special symbols such as "gamma," "integral," and "sub zero."

Proofs which involve diagrams form a special problem; however, we appeal again to the limitations of the human eye. Any diagram can be replaced by a diagram made up of line segments whose endpoints have rational coordinates, and which the human eye cannot distinguish from the original diagram. The line segment could then be replaced by an alphanumeric string giving the equations for these line segments.

There is a subtle point here. Are there proofs which the human mind can understand but which cannot be expressed in any formal logical system? If so, all bets are off and even Gödel's incompleteness theorem goes out the window. Certainly a proof with a diagram may be understandable, while a proof with equations for line segments would not be understandable. However, I wield a two-edged sword. On the one side, I argue the finiteness of the set of all distinct diagrams, by which I mean diagrams distinct to the human eye. So that even if we extend our symbol set to include diagrams, it remains finite (very large, but finite). On the other side I assert that either a proof involving perception of a diagram can be replaced (perhaps using, for example, the methods of analytic geometry) by a more formal argument, or else the human mind can manipulate infinite data in a single step, in which case it may be that nothing is impossible.

Similar arguments apply to proofs which use moving objects, musical notes, color, tactile sensations, taste, or smell.

2. A more serious problem arises with the fear that perhaps only long theorems have long proofs, coupled with the intuition that long theorems are less interesting than short theorems. It remains a possibility that all theorems with long proofs are uninteresting, though experience argues against this. (One could argue, similarly, that the speed limit set by the speed of light is unimportant, since all places far, far from home are uninteresting. One could argue, similarly, that the grapes are sour anyway.)

Because the virtue of an example is to be simple, let us consider again the class of theorems of the form "The integer p is prime," where the symbol p is replaced by a prime number to obtain a theorem. Does the statement "The integer p is not divisible by any number greater than one and less than or equal to the square root of p" constitute a proof of this theorem when the symbol p is replaced by the same prime? Of course not. If it did, then it would still be a proof if we replaced p by a composite number in both theorem and proof. A proof of this type must include a list, perhaps a computer printout, of divisions with remainders. Here the length of the proof grows much faster than the length of the theorem.

Similarly, the assertion "Every configuration in an unavoidable set U is reducible" is not in

itself a proof of the Four Color Theorem [1], though this assertion together with a "list" is a proof, even when the list exists only transiently in a computer's memory and is never committed to paper. Note that the computer memory, viewed serially, with each state followed sequentially by a new state marker and the next state, is still a finite string of alphanumeric characters.

It remains, however, an open and interesting question whether the ratio of the length of proofs to the length of theorems is unbounded.

3. The length of a proof depends strongly not only on the symbols in the proof but on information stored in the memory of a person reading the proof. Again we have two cases: either the information stored in the human mind is finite, in which case the proof of the main theorem at the beginning of this paper goes through exactly as before, or else the information stored in the human mind is infinite, in which case every theorem, including "Gödel undecidable" theorems, may have a short proof, i.e.,

It is intuitively obvious to the most casual observer.

A friend of mine has pointed out an analogy between length of proof and length of program—something computer scientists are very interested in. Given a theorem, you may need an entire book to prove the theorem to someone unacquainted with the subject; a few well-chosen phrases may suffice for a colleague whose research interests are close to your own; and the word "obviously" will suffice for someone who is convinced of your theorem already. Similarly, a very long string of bits is necessary to explain a program to a machine; if the computer already has a compiler in its memory, then a much shorter program in a high-level language will suffice; and if there is already a subroutine to do what you want to do, then your program may consist of a single word.

In any case, if the data already in memory (yours or the computer's) is finite, and the total amount of work to be done is infinite, then arbitrarily long strings are needed to do the job.

4. Question: does an analog system (for example, a voltmeter, an ammeter, or a device to vary voltage and resistance independently) in itself comprise a proof of the infinitely many theorems included in a multiplication table for the real numbers?

Answer: No. Why not?

5. Since human beings can read only a finite number of words a minute, and since human lifetimes are finite, it follows that there are true theorems whose proof the human mind can never comprehend, even though valid proofs exist.

There has been some grumbling about the use of a computer in Appel and Haken's proof of the four color theorem [1]. But given the existence of true theorems with proofs too long for human comprehension, the use of computers to prove theorems seems unavoidable. It is cold comfort that there are also true theorems whose shortest proof is too long for any computer to handle.

Research for this paper was supported by NSF Grant No. MCS 77-18723 A04.

References

1. K. Appel and W. Haken, Every planar map is four colorable, Bull. Amer. Math. Soc., 764 (1976) 711–712.

2. K. Gödel, Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter System I, Monatsh. Math. Phys., 38 (1931) 173–198.

MISCELLANEA

68. The mathematician knows everything—but nothing except that.

—Mirko Stojaković

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

HOW MANY DIFFERENT RINDS CAN YOU PEEL FROM A SEQUENCE?

FRITS GÖBEL
Department of Applied Mathematics, Technische Hogeschool Twente
Enschede, The Netherlands

Let $S = (s_1, s_2, \dots, s_n)$ be a sequence of natural numbers. We form the sequence $T = (t_1, t_2, \dots, t_n)$ by rearranging the elements of S as follows. For $i = 1, \dots, n$ the ith term of T is equal to the leftmost element or the rightmost element of the sequence S or what remains of it. After each assignment to T, the element just assigned is deleted from S. A sequence T derived from S in this manner is called a **rind** of S. The reason for the terminology is that T may be thought of as a stack of pieces of peel which have been successively shelled from one end or the other of S. More formally, a sequence $T = (t_1, t_2, \dots, t_n)$ is called a **rind** of $S = (s_1, s_2, \dots, s_n)$ if there is a sequence $(e_1, e_2, \dots, e_n)$ of zeros and ones such that T can be obtained from S by the algorithm:

```
L := 1; R := n;
for j := 1 step 1 until n do
if e_j = 1 then begin t_j := s_L; L := L + 1 end
else begin t_j := s_R; R := R - 1 end.
```

EXAMPLE. From $S = (1,2,3,1,4)$ we may obtain $T = (1,4,2,1,3)$ by taking alternately the leftmost and rightmost element of S. Or we may obtain $T' = (1,2,4,1,3)$ by choosing the elements of S in a different order. □

The number of different rinds of S is denoted by $\rho(S)$. It is easy to see that $\rho(S) \leqslant 2^{n-1}$ for all S of length n and that equality holds if and only if all terms of S are different. The determination of $\rho(S)$ for arbitrary S is an unexpectedly difficult problem.

DEFINITION. A sequence S is called a **grouped sequence** if $s_i = s_j$ with $i < j$ implies $s_i = s_k$ for all k with $i < k < j$.

CONJECTURE. *Let S be a sequence. Then the maximum of $\rho(S')$ over all rearrangements S' of S is attained for a grouped sequence.*

Partial result 1. If S contains exactly two different elements, say p ones and q twos, then $\rho(S') = \binom{p+q}{p}$ where S' is the sequence of p ones followed by q twos. Since $\binom{p+q}{p}$ is the number of rearrangements of p ones and q twos without any restriction, that value is indeed the maximum over all rearrangements of S.

Partial result 2. If the conjecture could be proved for sequences containing exactly 3 different elements, then we know *which* grouped sequence is optimal, namely, the one with the "rarest" element in the middle.

Partial result 3. If S is a grouped sequence of length n consisting of five groups of respective lengths a, b, c, d, e, then

$$\rho(S)=\binom{n}{a+b}+\binom{n}{d+e}+\binom{n}{a}+\binom{n}{e}-\binom{n-b}{a}-\binom{n-d}{e}-\binom{n-c}{a+b}.$$

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

AN "ARCHIMEDEAN" PARADOX

ANDREW SIMOSON
Department of Mathematics, King College, Bristol, TN 37620

Archimedes' Principle that between every pair of rational numbers an irrational one can be found and between every pair of irrational numbers a rational one can be found suggests that there are just as many rational numbers as irrational ones. The paradox is the failure of this obvious conjecture. Using somewhat the same frame of mind, one might expect that the real numbers can be uniformly partitioned in two, that is, that some mathematical comb exists which when drawn through the real numbers will remove them in such a way that in every subinterval of the real line "half" of the numbers are removed. It turns out that no such comb exists. However, such combs can be "approximated," and an example is given.

All work will be done on the open unit interval $I=(0,1)$. Recall that the *outer measure* of A, the natural generalization of length, denoted by $m(A)$, is $\inf\{\sum_{n=1}^{\infty} l(J_n)\colon A \subset \cup_{n=1}^{\infty} J_n\}$ where $\{J_n\}_{n=1}^{\infty}$ is any countable collection of open subintervals of I and $l(J_n)$ is the length of J_n. Since every open set U of real numbers is the union of a countable collection of pairwise disjoint open intervals, the outer measure of A is also equal to $\inf\{m(U)\colon A\subset U\}$ where U is any open subset of I. Outer measure is subadditive; that is, $m(\cup_{n=1}^{\infty} A_n)\leqslant \sum_{n=1}^{\infty} m(A_n)$ for any countable collection of sets $\{A_n\}_{n=1}^{\infty}$. We recall that a set A is *measurable* if for every set B we have $m(B)=m(A\cap B)+m(A'\cap B)$ where A' is the complement of A. Unions of finite collections of intervals and open sets are measurable. The measure of singleton sets is 0. Useful in the following is this result [3, p. 57]: For any set A and for any finite collection of disjoint measurable sets $\{C_n\}_{n=1}^{M}$,

$$m\left(A\cap\bigcup_{n=1}^{M} C_n\right)=\sum_{n=1}^{M} m(A\cap C_n). \qquad (*)$$

We define a subset A of I to be a *comb* if there exist numbers a and b with $0<a\leqslant b<1$ such that, for every nonempty open subinterval J of I, $a\leqslant m(A\cap J)/m(J)\leqslant b$. In particular, if we take a comb with $a=\frac{1}{2}=b$, then we are considering a set of numbers that could be described as containing exactly half of the numbers for any subinterval of I. It would seem plausible that some exotic manipulation of the Cantor set or some wild invocation of the axiom of choice ought to yield a comb. But no such scheme exists. The following demonstrates the paradox.

THEOREM. *No comb exists.*

Proof. Let A be a comb. Then (with $J=(0,1)$), $0<a\leqslant m(A)\leqslant b<1$. Let ε be a positive

number less than $(1/b - 1)m(A)$. Then choose U to be an open subset of I containing A such that $m(U) < m(A) + \varepsilon$. Write $U = \cup_{n=1}^{k} J_n$ where the J_n are pairwise disjoint nonempty subintervals of I and where k is either a positive integer or ∞. (Note that $m(J_n) > 0$ for all n.) Then

$$m(A) + \varepsilon > \sum_{n=1}^{k} m(J_n) = \sum_{n=1}^{k} (m(J_n)/m(A \cap J_n))m(A \cap J_n) \geq \sum_{n=1}^{k} (1/b)m(A \cap J_n)$$
$$= (1/b) \cdot \sum_{n=1}^{k} m(A \cap J_n) \geq (1/b)m(A),$$

which implies the contradiction that $\varepsilon \geq (1/b - 1)m(A)$.

Failing to find a comb but believing that most notions that can be imagined can be approximated to some degree, we define a subset A of I to be an *approximate comb* if $0 < m(A \cap J)/m(J) < 1$ for all nonempty open subintervals J of I. We say that a subset A of I has a *gap* J or a *block* J if $m(A \cap J) = 0$ or $m(A \cap J) = m(J)$, respectively, where J is a nonempty open subinterval of I. For example, the set of rational numbers in I has a gap of I while its complement in I has a block of I. Approximate combs are sets with neither gaps nor blocks. Define

$$\beta(A) = \sup\{m(A \cap J)/m(J) : J \text{ is a nonempty open subinterval of } I\}$$

and

$$\alpha(A) = \inf\{m(A \cap J)/m(J) : J \text{ is a nonempty open subinterval of } I\},$$

for any subset A of I. The proof of the theorem above yields the following corollary.

COROLLARY 1. *Let A be an approximate comb. Then $\beta(A) = 1$.*

COROLLARY 2. *Furthermore, if A is a measurable approximate comb, then $\alpha(A) = 0$.*

Proof. For all nonempty open subintervals J of I, $m(A' \cap J)/m(J) = 1 - m(A \cap J)/m(J)$ by definition of a measurable set, and we have $\alpha(A) \leq m(A \cap J)/m(J) < 1$ by definition of $\alpha(A)$. Hence $1 - \alpha(A) \geq m(A' \cap J)/m(J) > 0$ for all nonempty open subintervals J of I. Therefore $\beta(A') \leq 1 - \alpha(A)$. Since A' is an approximate comb, then $\beta(A') = 1$ by Corollary 1. Thus $\alpha(A) = 0$.

Hence approximate combs "almost" have either blocks or combs. An argument in [**1**, p. 274] and an exercise in [**4**, p. 189], utilizing the almost everywhere differentiability of monotone functions, give results similar to those of the two corollaries above, namely, that if A is a measurable subset of I with $0 < m(A) < 1$, then $\beta(A) = 1$ and $\alpha(A) = 0$.

An interesting question (which we shall leave open) is to determine if there exist subsets A of I such that $0 < \alpha(A) < 1$. In [**2**] an example of a subset C of I with $\alpha(C) \geq \frac{1}{2}$ is given; but it is conceivable for this set (or for any set C with a similar characteristic) that $m(C) = 1$ and hence that $\alpha(C) = 1$.

Before constructing an approximate comb we state some preliminary lemmas.

LEMMA 1. *Let a, b, c, d, r be nonnegative real numbers with b and d nonzero. If $a/b < r$ and $c/d \leq r$, then $(a + c)/(b + d) < r$.*

Proof. The addition of the inequalities $a < br$ and $c \leq dr$ yields $a + c < (b + d)r$, which gives the desired conclusion.

LEMMA 2. *Let J_1 and J_2 be nonempty open intervals with J_2 being a proper subset of J_1. Let A be a set of real numbers such that $0 < m(A \cap J_2)/m(J_2) < 1$. Then $0 < m(A \cap J_1)/m(J_1) < 1$.*

Proof. Since $m(A \cap J_2) > 0$ then $m(A \cap J_1)/m(J_1) > 0$. Let $K = J_1 \setminus J_2$. By hypothesis $m(K) > 0$. Let $a = m(A \cap J_2)$, $b = m(J_2)$, $c = m(A \cap K)$, $d = m(K)$, $r = 1$. By $(*)$ and

Lemma 1, $m(A \cap J_1)/m(J_1) < 1$.

Construction of an Approximate Comb. Let

$$D_n = \bigcup_{k=1}^{2^n} (k/2^n - 1/4^n, k/2^n) \quad \text{and} \quad E_n = \bigcup_{k=1}^{2^n - 1} (k/2^n, k/2^n + 1/4^n).$$

Define $A_n = D_n \setminus \cup_{j=n+1}^{\infty} E_j$. Define $A = \cup_{n=1}^{\infty} A_n$. We claim that A is an approximate comb.

Let J be a nonempty open subinterval of I. We shall show that $0 < m(A \cap J)/m(J) < 1$. Let p and q be integers greater than 1 with $B = ((p-1)/2^q, p/2^q) \subset J$. By Lemma 2, it will be sufficient to show that $0 < m(B \cap A) < m(B)$. To establish this inequality we make four notes.

Note 1. For all n, we have $0 < m(A_n) < 1/2^n$.

Proof. Observe that $m(D_n) = 1/2^n$ and that $m(E_n) < 1/2^n$. Since $D_n \subset A_n \cup (\cup_{j=n+1}^{\infty} E_j)$ and since outer measure is subadditive we have $m(D_n) \leqslant m(A_n) + \Sigma_{j=n+1}^{\infty} m(E_j)$. Hence

$$0 < m(D_n) - \sum_{j=n+1}^{\infty} m(E_j) \leqslant m(A_n) < 1/2^n.$$

Note 2. For any $n \leqslant q$, we have $(p/2^q - 1/4^q, p/2^q) \cap A_n \subset B \cap A_q$, since $B \cap A_q = (p/2^q - 1/4^q, p/2^q) \setminus \cup_{j=q+1}^{\infty} E_j$ and $\cup_{j=q+1}^{\infty} E_j \subset \cup_{j=n+1}^{\infty} E_j$.

Note 3. For any $n < q$, we have $((p-1)/2^q, (p-1)/2^q + 1/4^q) \cap A_n = \varnothing$, by definition of A_n and by the fact that p is greater than 1.

Note 4. For any $n \geqslant q$, we have $m(B \cap A_n) = (1/2^q)m(A_n) < (1/2^q)(1/2^n)$, using Note 1.

Hence we have the following inequalities.

$$0 < (1/2^q)\left\{m(D_q) - \sum_{j=q+1}^{\infty} m(E_j)\right\} \leqslant (1/2^q)m(A_q) = m(B \cap A_q) \leqslant m(B \cap A)$$

$$\leqslant m\left(((p-1)/2^q, p/2^q - 1/4^q) \cap \bigcup_{n=1}^{q-1} A_n\right)$$

$$+m\left(\{((p-1)/2^q, p/2^q - 1/4^q) \cap A_q\} \cup \left\{(p/2^q - 1/4^q, p/2^q) \cap \bigcup_{n=1}^{q} A_n\right\}\right)$$

$$+m\left(B \cap \bigcup_{n=q+1}^{\infty} A_n\right)$$

$$\leqslant m\left(((p-1)/2^q + 1/4^q, p/2^q - 1/4^q) \cap \bigcup_{n=1}^{q-1} A_n\right) \qquad \text{(by Note 3)}$$

$$+m(\{((p-1)/2^q, p/2^q - 1/4^q) \cap A_q\} \cup (B \cap A_q)) \qquad \text{(by Note 2)}$$

$$+ \sum_{n=q+1}^{\infty} m(B \cap A_n) \qquad \text{(by subadditivity)}$$

$$\leqslant m(((p-1)/2^q + 1/4^q, p/2^q - 1/4^q))$$

$$+m(B \cap A_q) + \sum_{n=q+1}^{\infty} m(B \cap A_n)$$

$$= 1/2^q - 2/4^q + \sum_{n=q}^{\infty} m(B \cap A_n)$$

[Continued on p. 125.]

$$\leqslant 1/2^q - 2/4^q + (1/2^q)\left\{\sum_{n=q}^{\infty} m(A_n)\right\} \qquad \text{(by Note 4)}$$

$$< 1/2^q - 2/4^q + (1/2^q)\left\{\sum_{n=q}^{\infty} 1/2^n\right\} = 1/2^q = m(B) \qquad \text{(by Note 1)}.$$

Therefore we have $0 < m(A \cap J)/m(J) < 1$ for all nonempty open subintervals J of I, and hence A is an approximate comb.

References

1. E. Hewitt and K. R. Stromberg, Real and Abstract Analysis, Springer-Verlag, 1965.
2. ______, Some examples of nonmeasurable sets, J. Austral. Math. Soc., 18 (1974) 236–238.
3. H. L. Royden, Real Analysis, Macmillan, New York, 1968.
4. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1974.

Who is the wise guru, and who are his disciples?

(Answer on page 129.)

CENTER SECTION
(Vol. 89, No. 2, Feb. 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota, 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P, L*.** Recommendations for a General Mathematical Sciences Program. MAA, 1981, 102 pp, $3.50 (P). [ISBN: 0-88385-437-6] The first major CUPM curriculum recommendation in nearly a decade, this long-awaited report suggests a more heterogeneous undergraduate major in mathematical science, substituting newer applications-oriented courses such as modelling, discrete methods, and computer science for the former CUPM-recommended advanced courses in topology, analysis, and algebra. The volume includes subpanel reports on calculus (shift to a two rather than a three semester basic sequence); core mathematics ("There is no longer a common body of pure mathematical information that every student should know."); computer science ("All mathematical science majors should be required to take the first [computer science] course, and strongly encouraged to take the second course); modelling ("Mathematical science majors should have substantial experience with mathematical modelling."); and statistics (the introductory course should "concentrate on data"). LAS

General, S, P, L**.** Notes on Rubik's Magic Cube. David Singmaster. Enslow Pub, 1981, vi + 73 pp, $5.95 (P); $9.95. [ISBN: 0-89490-043-9; 0-89490-057-9] A mathematician's exegesis of the notorious cube: subgroups, wreath products, analysis of algorithms. First official publication of the 1980 Preliminary Version (TR, October 1980). LAS

General, P, L.** Mathematical Reviews Annual Index-1980. Ed: J.L. Selfridge. AMS, 1981, $180 set (P). Part 1: Author Index A-L, 403 pp; Author Index M-Z, 404 pp; Part 2: Subject Index 00-58, 462 pp; Subject Index 60-94, 362 pp. Also identified as Index of Mathematical Papers, Volume 12. LAS

General, L*. A Dictionary of Named Effects and Laws in Chemistry, Physics and Mathematics, Fourth Edition. D.W.G. Ballentyne, D.R. Lovett. Chapman & Hall, 1980, viii + 346 pp, $20. [ISBN: 0-412-22380-5] About 2000 laws, formulas, units, definitions, and theorems from physical science and classical mathematics: 20 entries for Newton, 17 for Gauss, 14 for Cauchy, but none for Tychonoff or Borel. A handy reference to things that are otherwise hard to look up. (First Edition, TR, May 1971.) LAS

General, P, L. Scientific Strategies to Save Your Life: A Statistical Approach to Primary Prevention. Irwin D.J. Bross. Statistics, V. 35. Dekker, 1981, v + 259 pp, $19.75. [ISBN: 0-8247-1273-0] Case histories of the slow, difficult and frustrating task of applying scientific knowledge to effect changes in public health policy. The author passes on the hard lessons he has learned in the politics of health hazards. GHM

Elementary, S(13). Elementary Mathematics for Basic Chemistry and Physics. S.M. Gabbay. Basic Sci Preparation Center, 1980, v + 120 pp, $9.95 (P). [ISBN: 0-9604722-07] Review of percents, scientific notation, units of measurement, linear equations, functional relationships, graphs and logarithms. Relies on rules with no explanation or understanding. Several serious typographical errors. MW

Precalculus, T(13: 1). College Algebra. Bernard Kolman, Arnold Shapiro. Academic Pr, 1981, xvi + 456 pp, $17.95. [ISBN: 0-12-417884-7] Topics include: systems of equations, matrices and determinants, analytic geometry, the exponential and logarithm functions, mathematical induction. Progress checks, end-of-chapter summaries, and self-tests. Includes numerous warnings to the student, e.g., $\sqrt{a} + \sqrt{b} \neq \sqrt{a+b}$. Tightly written with attention to detail. JSG

Precalculus, S(13). Precalculus Mathematics in a Nutshell: Geometry, Algebra, Trigonometry. George F. Simmons. William Kaufmann, 1981, vii + 119 pp, $4.95 (P). [ISBN: 0-86576-009-8] A very brief summary of the mathematics needed for calculus. LLK

Precalculus, T(13: 1). Essentials of Precalculus Mathematics, Second Edition. Dennis T. Christy. Harper & Row, 1981, x + 598 pp, $18.95. [ISBN: 0-06-041303-4] Very similar to the First Edition (TR, March 1977) with the addition of harder algebra problems in Chapters 4-6, some new sections, and many more exercises at the end of chapters (rather than chapter tests). LLK

Education, P. Priorities in School Mathematics: Executive Summary of the PRISM Project. NCTM, 1981, 33 pp, free (P). [ISBN: 0-87353-174-4] Summary of an NSF supported survey of over 12,000 persons interested in school mathematics (teachers, administrators, school board and PTA presidents) on issues raised in NCTM's An Agenda for Action (TR, December 1980). Sample insights: secondary teachers and laymen assigned higher priority to drill on basics in elementary school than to solving word problems; computer literacy is the favored choice for developing new materials for grades 7-12; there is very little support for NCTM's recommendations to reduce emphasis on potentially obsolete calculation algorithms (e.g., long division) or for introducing such "modern" topics as probability and statistics. LAS

Education, S(16-17), P, L. Mathematical Discovery: On Understanding, Learning, and Teaching Problem Solving, Combined Edition. George Polya. Wiley, 1981, xxv + 220 pp, $18.95 (P). [ISBN: 0-471-08975-3] Reprint in a single paperback volume of the corrected edition of the two-volume work first published in 1962 and 1965, with a preface by Peter Hilton, an updated bibliography (by Gerald Alexanderson) and an improved index (by Jean Pederson) which references both this volume and Polya's earlier works How to Solve It and Mathematics and Plausible Reasoning. This classic text on problem solving--written two decades before the current emphasis came into vogue--is still ripe with strategy, insight, heuristics, and hints for teachers. LAS

Education, T(17-18: 1), P, L. Curriculum Development in Mathematics. Geoffrey Howson, Christine Keitel, Jeremy Kilpatrick. Cambridge U Pr, 1981, viii + 288 pp, $49.50. [ISBN: 0-521-23767-X] A review of the rise and fall of mathematics curriculum reforms of the 1960's and 1970's, especially (but not exclusively) in the United States and the United Kingdom. Historical and educational contexts, patterns of project management, curriculum theory, and evaluation schemes combine to form a theoretical basis for future curriculum efforts. Includes discussion questions and projects for graduate students. LAS

Education, S(16), P. Mathematics Education Research: Implications for the 80's. Ed: Elizabeth Fennema. Assoc for Supervision & Curr Develop, 1981, x + 173 pp, $6.75 (P). [ISBN: 0-87120-107-0] Synthesis of research in important areas, including curriculum, problem solving, sex factors, calculators and computers. Implications for teaching and suggestions for further research. Valuable for in-service teachers and mathematics educators; could also be an understandable introduction to mathematics education research for pre-service teachers. MW

History, S(16-18), P, L*. A History of the Calculus of Variations from the 17th through the 19th Century. Herman H. Goldstine. Stud. in History of Math. and Phy. Sci., No. 5. Springer-Verlag, 1980, xviii + 410 pp, $48. [ISBN: 0-387-90521-9] A review of key papers from Fermat (principle of least time), Newton (motion in a resisting medium), and John Bernoulli (brachystochrone problem) to Hilbert, Kneser, Carathéodory and Hahn at the beginning of the twentieth century. An excellent resource, both for those interested in history and for teachers and students of calculus of variations. LAS

History, P, L. American Mathematical Heritage: Algebra and Applied Mathematics. Ed: Dalton Tarwater, et al. Math. Ser., No. 13. Texas Tech U, 1981, 127 pp, (P). Four papers on the history of algebra (S. MacLane, W. Feit, L. Fuchs, O. Taussky) and four on the history of applied mathematics (P. Lax, S. Ulam, R. Sellman, R. Kalman) from the third and fourth (1975, 1976) Texas Conferences on the "American Mathematical Heritage." (For other volumes in this series see Graduate Studies, No. 13, 1976, Men and Institutions in American Mathematics (TR, February 1977); and On the History of Statistics and Probability, Dekker, 1976.) LAS

History, P, L*. Code of the Quipu: A Study in Media, Mathematics, and Culture. Marcia Ascher, Robert Ascher. U of Michigan Pr, 1981, vii + 166 pp, $8.95 (P); $18.95. [ISBN: 0-472-06325-1; 0-472-09325-8] The Incas of Peru, who possessed no writing, use elaborately knotted colored cotton or wool cord, called quipus, to record and transmit numerical information. The authors of this study --an anthropologist and a mathematician--have decoded these quipus, thereby revealing a sophisticated base-ten bookkeeping system complete with cross-categorization and hierarchical structure. "With pieces of string the Incas developed a form of recording that forces reconsideration of writing as we generally understand that term." A fascinating glimpse into the development of mathematics in a culture quite remote from contemporary Western and Arabic influences. LAS

History, P*, L.** Gauss, A Biographical Study. W.K. Bühler. Springer-Verlag, 1981, viii + 208 pp, $16.80. [ISBN: 0-387-10662-6] A slim but rich "guidebook" for contemporary scientists and mathematicians, providing clues to connections with our own age rather than "antiquarian" reconstruction of past detail. Bühler effectively intersperses social and historical commentary with brief descriptions of Gauss's scientific and mathematical work, from least squares ("the most important witness [for Gauss] to the connection between mathematics and nature") to magnetism. Appendices contain a commentary and index to Gauss's collected works, a guide to the secondary literature, notes to the text, and a bibliography of other Gauss biographies. LAS

History, P, L. Marston Morse: Selected Papers. Ed: Raoul Bott. Springer-Verlag, 1981, xlvi + 882 pp, $36. [ISBN: 0-387-90532-4] 35 of Morse's 176 papers, preceded by Bott's biography of Morse (reprinted from the Bull. Amer. Math. Soc.). Concludes with a complete Morse bibliography. LAS

History, S, P, L.** From ENIAC to UNIVAC: An Appraisal of the Eckert-Mauchly Computers. Nancy Stern. Digital Pr, 1981, ix + 286 pp, $21. [ISBN: 0-932376-14-2] A fascinating and thoroughly documented historical analysis of scientific, commercial and policy issues involved in the evolution of

EDVAC, the first stored-program computer, and UNIVAC, the first commercial computer. The record of contention and patent conflict between academic and commercial interests is a vivid object lesson for today's bioengineering debates. Includes as a lengthy appendix the first complete publication of von Neumann's 1945 seminal "Report on EDVAC" which articulates computer architecture. Author Stern argues that credit for conceiving the stored-program concept should go to the Eckhert-Mauchly group that designed ENIAC and EDVAC, and that von Neumann's role was merely to recognize the potential of this idea and publicize it. LAS

Foundations, T(18), P. The Lambda Calculus: Its Syntax and Semantics. H.P. Barendregt. Stud. in Logic and Found. of Math., V. 103. Elsevier North Holland, 1981, xiv + 615 pp, $109.75. [ISBN: 0-444-85490-8] Detailed and well organized exposition of the pure lambda calculus (roughly: "theory of symbolic computation," akin to recursion theory), broken into four parts on Conversion, Reduction, Theories, and Models. Modest exercise sets. Though the applications to computer science are not discussed, the lambda calculus confronts in pure form many of the problems involved in programming languages. May become the standard reference work. GHM

Foundations, P, L. Frege and the Philosophy of Mathematics. Michael D. Resnik. Cornell U Pr, 1980, 244 pp, $15. [ISBN: 0-8014-1293-5] Frege's profound contributions to the philosophy of mathematics are described by presenting Frege's incisive criticisms of the views of his contemporaries (Psychologism, Formalism, Deductivism and Mill's Empiricism). Fregean criticisms of modern derivative views are also woven into the discussion. The final third of the book expounds and criticizes Frege's positive doctrines on logic, epistemology and the nature of number. GHM

Foundations, P, L. Mechanism, Mentalism, and Metamathematics: An Essay on Finitism. Judson Chambers Webb. Synthese Library, V. 137. D Reidel Pub, 1980, xiii + 277 pp, $28.95. [ISBN: 90-277-1046-5] An essay on the psychological and philosophical significance of various metamathematical theorems. Argues that, contrary to widely held opinions, both Formalism and Mechanism are supported and strengthened by the limitative theorems of Gödel and Church. GHM

Foundations, P. Decidability and Boolean Representations. Stanley Burris, Ralph McKenzie. Memoirs No. 246. AMS, 1981, vii + 106 pp, $6.40 (P). Part I contains quite general characterizations of decidable algebraic varieties in terms of decompositions of varieties. Part II studies consequences of sheaf representations of varieties (particularly as Boolean powers) for their decomposability. GHM

Linear Algebra, T(14: 1). Linear Algebra with Applications. Steven J. Leon. Macmillan, 1980, xii + 338 pp, $17.95. [ISBN: 0-02-369870-5] A good linear algebra text with few applications. Has starred sections covering such topics as complex matrices, quadratic forms, and numerical linear algebra. LLK

Linear Algebra, S(15-16). Einführung in die Lineare Algebra and Geometrie, Teil 2. Johann Cigler. Manzsche Verlag, 1976, 118 pp, (P). [ISBN: 3-214-00020-9] A fairly elementary but modern introduction to finite-dimensional vector spaces over a field. No exercises or index. JD-B

Linear Algebra, T(17-18: 1), S, P, L. Finite and Infinite Dimensional Linear Spaces: A Comparative Study in Algebraic and Analytic Settings. Richard D. Järvinen. Marcel Dekker, 1981, xiv + 168 pp, $27.25. [ISBN: 0-8247-1172-6] An interesting comparative study of linear spaces, taking the basic notions for finite dimensions and extending or modifying them appropriately for infinite dimensions, pointing out potential pitfalls along the way. The essential preliminaries are followed by a chapter on infinite systems of equations and the Fredholm theory and a chapter on topological (mostly normed) linear spaces. The last chapter is a list of some current research problems. Could be used as a graduate text; exercises, index, bibliography. JS

Algebra, P. Lecture Notes in Mathematics-848: Algebra, Carbondale 1980. Ed: R.K. Amayo. Springer-Verlag, 1981, vi + 298 pp, $19.50 (P). [ISBN: 0-387-10573-5] Proceedings of the conference held at Southern Illinois University, April 18-19, 1980. JAS

Algebra, P. Lecture Notes in Mathematics-831: Representation Theory I. Ed: V. Dlab, P. Gabriel. Springer-Verlag, 1980, xiv + 373 pp, $22 (P). [ISBN: 0-387-10263-9] Five series of lectures followed by a 50-page bibliography on developments during the period 1974-79 following the First International Conference on Representations of Algebras (Springer Lecture Notes #488): P. Gabriel on Auslander-Reiter sequences, J.E. Humphreys on highest weight modules, C.M. Ringel on the Brauer-Thrall conjectures and on tame algebras, and V.A. Rojter on representation of bimodules with a coalgebra structure (BOCS). LAS

Algebra, P. Groupes Abéliens sans Torsion. Khalid Benabdallah. Presses U Montreal, 1981, 180 pp, (P). [ISBN: 2-7606-0545-0] Presents an exposition of recent trends first by discussing groups of rank 1, of rank 2, of p-rank 1, and quasi-endomorphisms, and secondly by listing results in many papers published between 1972 and 1978. SG

Algebra, T(17-18), S, P, L. Universal Algebra. Paul M. Cohn. Math. and Its Appl., V. 6. D. Reidel Pub, 1981, xv + 412 pp, $19.50 (P); $44.50. [ISBN: 90-277-1254-9; 90-277-1213-1] A complete revision of the 1965 edition (TR, April 1967) including the addition of four new chapters. The many good exercises and the intention to avoid maximum generality make it suitable as a text for beginning graduate students. JG

Complex Analysis, P. Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference. Ed: Irwin Kra, Bernard Maskit. Annals of Math. Stud., No. 97. Princeton U Pr, 1981, ix + 517 pp, $25; $9.50 (P).

Differential Equations, T(16-18: 1, 2), S, P, L. Catastrophe Theory for Scientists and Engineers. Robert Gilmore. Wiley, 1981, xvii + 666 pp, $45.95. [ISBN: 0-471-05064-4] A comprehensive guide to the equations and geometry of catastrophe theory, "impedance matched" to the needs of engineers and physical scientists. Seven chapters on applications (thermodynamics, mechanics, aerodynamics, caustics, canonical forms, quantum mechanics), and four on extensions beyond elementary catastrophe theory are sandwiched between nine chapters ("for pedestrians") introducing the elementary catastrophes and three chapters ("for joggers") outlining Thom's theory. LAS

Differential Equations, T(17-18: 1), P. Ordinary Differential Equations: Stability and Periodic Solutions. N. Rouche, J. Mawhin. Trans: R.E. Gaines. Pitman Pub, 1980, ix + 260 pp, $49.95. [ISBN: 0-273-08419-4] An advanced text which illustrates the qualitative and analytical approaches to differential equations by means of the study of the stability and periodicity of solutions. The original French edition was published in 1973. AO

Differential Equations, P. Qualitative Analysis of the Periodically Forced Relaxation Oscillations. Mark Levi. Memoirs No. 244. AMS, 1981, vi + 147 pp, $8.40 (P). The author applies results and methods of dynamical systems to the study of second-order differential equations of van der Pol type with periodic forcing. Qualitative behavior (e.g., how many stable periodic solutions) of solutions for various values of the amplitude parameter is analyzed by deriving a geometrical representation for the system. PZ

Differential Equations, P. Lecture Notes in Mathematics-858: Regular Boundary Value Problems Associated with Pairs of Ordinary Differential Expressions. Earl A. Coddington, Hendrik S.V. de Snoo. Springer-Verlag, 1981, 225 pp, $14 (P). [ISBN: 0-387-10706-1] A detailed investigation of eigenvalue problems associated with pairs of ordinary differential operators when the coefficients of the operators are nice on a closed bounded interval and one of them is positive definite. AO

Differential Equations, S(14). A Book of Problems in Ordinary Differential Equations. M.L. Krasnov, A.I. Kiselyov, G.I. Makarenko. Trans: Vldimir Shokurov. MIR Pub, 1981, 332 pp, $6.80 (P). English translation of the third Russian edition. Contains worked-out examples and 967 mostly routine exercises of standard material through systems and stability theory. Easy to read. Excellent diagrams. Answers to almost all exercises. Not many surprises, but recommended as a problem source for the teacher. JK

Differential Equations, T(17-18: 1), P. Singular Systems of Differential Equations. S.L. Campbell. Research Notes in Math., No. 40. Fearon Pitman Pub, 1980, 176 pp, $17.95 (P). [ISBN: 0-8224-8438-2] On systems of the form $A\underline{x}' + B\underline{x} = \underline{f}$ and the discrete analog $A\underline{x}_{k+1} + B\underline{x}_k = \underline{f}_k$, where A is singular. Emphasis is on deriving closed forms for solutions. For postgraduate students in pure and applied mathematics, electrical engineering and systems science. Linear algebra and functional analysis are required. Examples, some exercises for earlier chapters and an admittedly incomplete bibliography. JK

Numerical Analysis, P. Applications of Number Theory to Numerical Analysis. Hua Loo Keng, Wang Yuan. Springer-Verlag, 1981, viii + 241 pp, $39. [ISBN: 0-387-10382-1] A presentation of approximation techniques based on results of algebraic number theory. Applications are given to numerical integration of periodic functions, interpolation, integral and differential equations. SG

Functional Analysis, S(17-18), P. Theory of Linear Operators in Hilbert Space. N.I. Akhiezer, I.M. Glazman. Trans: E.R. Dawson, W.N. Everitt. Pitman Pub, 1981. Volume I, xxxii + 312 pp, $11.50 [ISBN: 0-273-08495-X]; Volume II, xxxii + 239 pp, $16.50. [ISBN: 0-273-08495-8] Translation of the third Russian edition (1977) of this well-known reference. This new edition incorporates a significant amount of new material. AO

Analysis, P. Polylogarithms and Associated Functions. Leonard Lewin. Elsevier North Holland, 1981, xvii + 359 pp, $54.95. [ISBN: 0-444-00550-1] Revised and updated version of an earlier work by the author, Dilogarithms and Associated Functions. It is probably the most comprehensive reference on these functions currently available. AO

Analysis, T(14-15: 1). Methods of Applied Mathematics. N.M. Queen. Thomas Nelson & Sons, 1980, vii + 231 pp, $4.50 (P). [ISBN: 0-17-771121-3] Traditional material on methods of applied mathematics for sophomores and juniors majoring in the physical sciences. Emphasis on methods rather than on theory. Chapter coverage somewhat flexible. Some real analysis background required. Necessary linear algebra reviewed. Problems with answers where appropriate. Non-encyclopedic. Of suitable length for one-semester course. Worth a look if price is a factor in adoption decision. JK

Analysis, S(16-17), L. Mathematics for Engineers and Scientists, Second Edition. Alan Jeffrey. Thomas Nelson & Sons, 1979, x + 734 pp, $6.95 (P). [ISBN: 0-17-771605-3] An extensive revision of the original edition (TR, April 1970), this is an unusually broad and concise survey of most of the undergraduate program in analysis, beginning with number systems then moving briskly through calculus, linear algebra, vector analysis, differential equations, numerical analysis, and probability. Not likely to be used as a text (it is really more of a handbook), but it does include some proofs, examples, and exercises. Index, answers. JS

Algebraic Geometry, P. Lecture Notes in Mathematics-847: Vector Fields and Other Vector Bundle Morphisms--A Singularity Approach. Ulrich Koschorke. Springer-Verlag, 1981, iv + 304 pp, $18 (P). [ISBN: 0-387-10572-7]

Algebraic Geometry, P. Journées de géométrie algébrique d'Angers (1979): Variétés de petite dimension. Ed: Arnaud Beauville. Sijthoff & Noordhoff, 1980, xii + 323 pp. [ISBN: 90-286-0500-2] Conference proceedings containing 17 papers on vector bundles, periods of algebraic varieties, surfaces, and 3-dimensional varieties. SG

Algebraic Geometry, T(18), P. Basic Theory of Algebraic Groups and Lie Algebras. Gerhard P. Hochschild. Grad. Texts in Math., V. 75. Springer-Verlag, 1981, viii + 267 pp, $32. [ISBN: 0-387-90541-3] A second-year graduate text whose aim is to exhibit "basic algebra in action" rather than to develop the subject matter in extreme depth. Develops basic Lie algebra theory through semisimplicity and algebraic groups through Borel subgroups. SG

Differential Geometry, P. Existence and Regularity of Minimal Surfaces on Riemannian Manifolds. Jon T. Pitts. Princeton U Pr, 1981, 329 pp, $12 (P). Presents the work of the author establishing the existence and properties of regular minimal submanifolds of a Riemannian manifold. SG

Geometry, S*(15-18), P*, L*. Turtle Geometry: The Computer as a Medium for Exploring Mathematics. Harold Abelson, Andrea A. diSessa. MIT Pr, 1981, xx + 477 pp, $20. [ISBN: 0-262-01063-1] Turtle is a mathematical creature tht crawls across computer screens, tracing lines, curves and patterns that illustrate great ideas in geometry: vectors, curvature, space-filling curves, topology of curves, even curved space and relativity. Part of Seymour Papert's Logo project at M.I.T., turtle geometry is an innovative means of enriching mathematics learning from kindergarten through graduate school. This book introduces both geometry as discovered by turtle--or by turtle's human controller--and the principles of programming--data structures, definitions, recursion--required to implement turtle geometry. Contains numerous exercises (with hints and answers in the back) and two appendices suggesting how turtle geometry can be implemented in standard languages (Basic, Pascal). LAS

Geometry, T(16-17: 1, 2), S, L. Foundations of Three-Dimensional Euclidean Geometry. Izu Vaisman. Pure and Appl. Math., V. 56. Dekker, 1980, ix + 268 pp, $35. [ISBN: 0-8247-6901-0] Carefully written, modern, rigorous axiomatic construction of three-dimensional geometry. Suitable for the mathematically mature upperclassman or beginning graduate student in course on the foundations of geometry. Problems, with hints for their solutions, complement the text proper. Should be of interest to logicians and philosophers. JK

Topology, P. Ordinal Invariants in Topology. V. Kannan. Memoirs No. 245. AMS, 1981, iv + 164 pp, $9.60 (P). A map θ from category A to category B is a B-invariant in A if A_1 isomorphic to A_2 in A implies $\theta(A_1)$ isomorphic to $\theta(A_2)$ in B. This paper examines the situation when A is a subcategory of TOP and B is the category of ordinal numbers. JG

Optimization, P. Lecture Notes in Control and Information Sciences-31: Projection Methods in Constrained Optimisation and Applications to Optimal Policy Decisions. Berc Rustem. Springer-Verlag, 1981, xv + 315 pp, $17.40 (P). [ISBN: 0-387-10646-4] Development and review of projection techniques for quadratic and other non-linear programming problems. Applications to specification of suitable objective function for policy decisions. JSG

Probability, T(16-17: 1), S, L. Point Processes. D.R. Cox, Valerie Isham. Chapman & Hall, 1980, viii + 188 pp, $16.50. [ISBN: 0-412-21910-7] Puts emphasis on recently developed results and methods that are directly useful in applications of point processes. FLW

Probability, T(11-14: 1), S, L. The Mathematics of Games and Gambling. Edward W. Packel. New Math. Lib., V. 28. MAA, 1981, x + 141 pp, $8.75 (P). [ISBN: 0-88385-628-X] A beautiful blend of game lore and elementary probability, introducing expectation, permutations, binomial distribution, game theory and power indices via examples from backgammon, craps, poker, blackjack and lotteries. Developed as a text for an experimental freshman seminar, it eschews general theory in favor of specific examples, and uses only elementary high school algebra. LAS

Probability, T(16-17: 1, 2). Applied Probability. Frank A. Haight. Math. Concepts and Methods in Sci. and Eng., V. 23. Plenum Pr, 1981, xi + 290 pp, $35. [ISBN: 0-306-40699-3] Nonstandard introduction, as evidenced by the chapter headings: discrete probability, conditional probability, Markov chains, continuous probability distributions (the normal distribution is never mentioned), continuous time processes, and the theory of queues. Concerned with the construction and analysis of probability models most appropriate for operations research. RSK

Statistics, T(13-15: 1, 2), S, L**.** Elements of Statistical Analysis. Hans W. Gottinger. Walter de Gruyter, 1980, 244 pp, $35.25. [ISBN: 3-11-007169-X] Foundations of probability and utility, Bernoulli trials, Gaussian analysis, Bayesian estimation, likelihood functions, sufficient statistics, information theory, decision theory, sequential analysis, optimal stopping. Most topics require no calculus. FLW

Statistics, P. Statistical Performance of Location Estimators. C.A.J. Klaassen. Math. Centre Tracts, No. 133. Math Centrum, 1981, 100 pp, Dfl. 12.60 (P). [ISBN: 90-6196-209-9]

Statistics, T(13-14: 1, 2), S, L. Statistical Methods for Business and Economics, Revised Edition. Roger C. Pfaffenberger, James H. Patterson. Richard D Irwin Pub, 1981, xvii + 828 pp, $23.95. [ISBN: 0-256-02350-6] Presupposes no college mathematics. Standard topics plus multiple regression, time series, forecasting, non-parametric statistics, and decision theory. FLW

Statistics, T(16-18: 1), S. Simultaneous Statistical Inference, Second Edition. Rupert G. Miller, Jr. Springer-Verlag, 1981, xvi + 299 pp, $22. [ISBN: 0-387-90548-0] The original 1966 McGraw-Hill edition (TR, October 1967) along with the author's 1977 review articles on multiple comparisons. FLW

Statistics, T(16-18: 1), S, P, L. Sampling From a Finite Population. Jaroslav Hájek. Ed: Václav Dupac. Statistics, V. 37. Marcel Dekker, 1981, v + 247 pp, $27.50. [ISBN 0-8247-1291-1] Bayesian and robust approaches, methods of sampling, methods of estimation. No exercises. FLW

Statistics, T?(13-16: 1), S. Statistics for Biologists. D.J. Finney. Chapman & Hall, 1980, viii + 165 pp, $9.95 (P). [ISBN: 0-412-21540-3] Introduction to the principles and objectives of statistics, including some techniques. Contains solutions to all exercises. RSK

Statistics, P. Proceedings of the Twenty-Sixth Conference on the Design of Experiments. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC), 1981, xviii + 587 pp, (P). Papers presented at an October 1980 conference held at New Mexico State University, Las Cruces, New Mexico. LAS

Computer Programming, P, L*. 6809 Assembly Language Programming. Lance A. Leventhal. Osborne/McGraw-Hill, 1981, xxv + 529 pp, $12.50 (P). [ISBN:0-931988-35-7] Leventhal appears to have a formula for producing programming manuals. If so, it's a good formula and has produced another clear and thorough manual for the serious programmer. JAS

Computer Programming, S(14-15), P. Z8000 Assembly Language Programming. Lance A. Leventhal, Adam Osborne, Chuck Collins. Osborne/McGraw-Hill, 1980, xix + 898 pp, $12.50 (P). [ISBN: 0-931988-36-5] An introduction to assembly language programming and a detailed description of the instruction set of the 16-bit Z8000 microprocessor. A large number of sample programs are included. AO

Computer Programming, T(13: 1). FORTRAN: Getting Started. William S. Davis. Addison-Wesley, 1981, vii + 168 pp, $5.95 (P). [ISBN: 0-201-03104-3]; BASIC: Getting Started, 1981, vii + 152 pp, $5.95 (P). [ISBN: 0-201-03258-9] Both books start from the premise that this is a first programming language. Most of the same exercises are used in each, the discussion is adjusted to fit the language. Attention to detail is good--a beginner really could "get started" with this approach. LLK

Computer Programming, T(14: 1), S. The FORTRAN Cookbook. Thomas P. Dence. TAB Books, 1980, 334 pp, $8.95 (P). [ISBN: 0-8306-1187-8] This is not a text on Fortran. It is a book on problem solving using Fortran. The problems vary widely, from Egyptian fractions to Riemann-Stieltjes integrals to game theory. Has been used in a course entitled "Special Topics in Mathematics." LLK

Computer Programming, S*, P*, L*. The Elements of Programming Style, Second Edition. Brian W. Kernighan, P.J. Plauger. McGraw-Hill, 1978, xii + 168 pp, $6.95 (P). [ISBN: 0-07-034207-5] Extensive revision of this 1974 classic, the programmer's Strunk and White. Pithy rules ("Don't stop at one bug." "10.0 times 0.1 is hardly ever 1.0." "Make it right before you make it faster.") are illustrated with real examples from programming textbooks (including some from the earlier version of Elements itself). LAS

Computer Programming, P, L. Microsoft FORTRAN. Paul M. Chirlian. Dilithium Pr, 1981, ix + 333 pp, $14.95 (P). [ISBN: 0-918398-46-0] Rather complete coverage of the language and typical small computer operating environments for the complete novice. Incorporates structured programming techniques, flowcharts, algorithm development, debugging and program documentation. GHM

Computer Science, P, L. Computer Graphics. Infotech State of the Art Report, Ser. 8, No. 5. Infotech Ltd, 1980. Analysis, iii + 247 pp; Invited Papers, ii + 301 pp, $80 set. [ISBN: 8553-9660-1] An exposition of key issues in computer graphics (hardware, software, applications) followed by 19 commissioned papers on a wide variety of topics. Includes a useful annotated bibliography of references and resources in computer graphics. LAS

Computer Science, T*(14-16: 1, 2), S, L. Introduction to Computer Organization. Ivan Tomek. Computer Sci Pr, 1981, xvii + 456 pp, $21.95. [ISBN: 0-914894-08-0] A well-conceived text that treats computer hardware organization without getting into details about electronics. It covers all topics recommended for CS4 of ACM curriculum '78. With exercises and a workbook (unavailable to the reviewer). JL

Computer Science, S(15-18), P. The Minicomputer in On-Line Systems: Small Computers in Terminal-Based Systems and Distributed Processing Networks. Martin Healey, David Hebditch. Winthrop Pub, 1981, xviii + 334 pp, $22.95. [ISBN: 0-87626-579-4] A handbook intended for data processing professionals that provides an overview of the subject. JL

Applications (Biology), P, L. Lecture Notes in Biomathematics-40: Renewable Resource Management. Ed: Thomas L. Vincent, Janislaw M. Skowronski. Springer-Verlag, 1981, xii + 236 pp, $15 (P). [ISBN:

0-387-10566-2] Proceedings of a January 1980 workshop at the University of Canterbury on models available to managers of fisheries and other renewable ecosystems. Main theme: simple models (e.g., Lotka-Volterra and logistic) do not work in most realistic management situations. The 16 papers in this volume offer a variety of more realistic (and more complex) models. LAS

Applications (Engineering), T(15-17: 1, 2), S, P, L. Advanced Mathematics for Engineers. Wilfred Kaplan. Addison-Wesley, 1981, xix + 929 pp, $22.95. [ISBN: 0-201-03773-4] A comprehensive leisurely survey of much of the traditional advanced calculus and differential equations needed for the physical sciences and engineering. Similar in spirit and content to the author's Advanced Calculus but with several contemporary additions such as generalized functions, the fast Fourier transform, and the finite-element method. Very readable, with many examples, exercises. Index. answers. JS

Applications (Engineering), T(17-18: 1, 2), S, P, L. Applications of Functional Analysis in Engineering. J.L. Nowinski. Math. Concepts and Methods in Sci. and Eng., V. 22. Plenum Pr, 1981, xv + 304 pp, $37.50. [ISBN: 0-306-40693-4] With the intent of making accessible to engineers some of the techniques of functional analysis, the first ten chapters proceed from elementary linear algebra on to Hilbert space and various function spaces. The final six chapters consist of applications, including the method of the hypercircle, orthogonal projections, Rayleigh-Ritz methods, and distributions. Exercises, answers, index, bibliography. JS

Applications (Modelling), S(16-18), P. Energy Policy Planning. Ed: B.A. Bayraktar, et al. Plenum Pr, 1981, ix + 467 pp, $49.50. [ISBN: 0-306-40631-4] Keynote address, policy maker's views, working group reports and 21 invited papers from a November 1979 NATO Advanced Research Institute at Brookhaven National Laboratory. The Institute's theme was to examine the role of mathematical modelling and systems science in current energy policy analysis. A rich compendium of contemporary issues in mathematical modelling; a good resource for a seminar in modelling. LAS

Applications (Physics), S*, L. Particles: An Introduction to Particle Physics. Michael Chester. New Amer Lib, 1978, 166 pp, $2.50 (P). A breezy survey of atomic models, from Democritus to quarks, relating new evidence to new models in a clear, compelling scientific mystery tale. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Seaway Section

The Seaway Section held its fall meeting at S.U.N.Y. College at Brockport on November 6 and 7, 1981. There was a registered attendance of 85 mathematicians plus 10 student participants from Brockport.

Invited Address:

"The End of the World on a Microcomputer," by Kimyong Kim of S.U.N.Y. College at Brockport.

Short Presentations:

"Data Encryption Based on the Factoring of Large Numbers," by Jack Hollingsworth, Rochester Institute of Technology.
"Koebe Function Representations of Univalent Functions," by Guy Johnson, Syracuse University.
"On Mappings of the Euclidean Plane and its Subplanes," by Howard E. Bell, Brock University.
"Relational Data Bases," by Patricia A. Woodworth, Ithaca College.
"A Case Study of Highway Fatality Rates and the 55 MPH Speed Limit," by David Farnsworth, Eisenhower College.
"The 22nd International Mathematical Olympiad," by Nura Turner, S.U.N.Y. at Albany.

"Translates of Averages," by Robert W. Sloan, Alfred University.
"A New Look at Analysis--Infinite Arrays in APL," by R.W.W. Taylor, Rochester Institute of Technology.
"Analytic Geometry Revisited or How to Keep Busy Though Retired," by Emmet C. Stopher, S.U.C. at Oswego.
"Recent Developments on Summability Invariants," by S-C Chang, Brock University.

Panel Discussion:

"What is in a Computer Science Curriculum?" by Jack Hollingsworth, R.I.T., Dennis Martin, Brockport, and Patricia Woodworth, Ithaca College.

Concurrently with these sessions the staff of S.U.N.Y. at Brockport made available a microcomputer display.

North Central Section

The 1918 fall meeting of the North Central Section was held at Bemidji State University, Bemidji, Minnesota on October 23 and 24, 1981.

Invited Addresses:

"Seminar in a Convex Container," by Wayne Roberts, Macalester College.
"The Role of Combinatorics," by N.S. Mendelsohn, University of Manitoba.

Short Presentations:

"Counting Hands in Pinochle," by Sabra Anderson, University of Minnesota at Duluth.
"A Tridiagonal Matrix and Perfect Numbers," by Gerald Bergum, South Dakota State University.
"Microcomputing at Saint Olaf College," by Arthur Seebach and Lynn Steen, St. Olaf College.
"Critters and More Critters: A Leslie Model with Density Dependence," by Charles Caswell, Carleton College (student).
"Injective Envelopes of Modules Over Generalized Noetherian Rings," by Francis Hannick, Mankato State University.
"Problem Solving in the Undergraduate Curriculum," by Wojciech Komornicki, Hamline University.
"A Rapidly Growing Function," by George Mills, Carleton College.
"Excellent Triangles," by Keith Pierce, University of Minnesota at Duluth.

Special features included a magic show by Professor Mendelsohn, a demonstration of microcomputer graphics by the Saint Olaf College contingent, and a showing of Mathematics at Work in Society videotapes.

Maryland-District of Columbia-Virginia Section

The fall meeting of the Maryland-District of Columbia-Virginia Section was held on November 13-14, 1981 at George Washington University, Washington, D.C. Eighty registrants attended the meeting.

Invited Lectures:

"Implications of Prime 80 on Applied Mathematics," by Alan Tucker, SUNY at Stony Brook.
"The Teaching of Mathematical Problem Solving," by Carolyn Maher, Rutgers University.

Short Presentations:

"The Game of Logic," by Gail Kaplan, U.S. Naval Academy.
"Peg Jump Games," by William Wardlaw, U.S. Naval Academy.
"How Many Lines in Space Intersect 4 Given Lines," by James Stormes, U.S. Naval Academy.
"Research into Concepts of Remediation," by Gloria Gilmer, National Institute of Education.
"The Status of Mathematics at N.S.F.," by Alvin Thaler, National Science Foundation.
"Integer Matrices with Integer Inverses," by John R. Hanson, James Madison University.
"The Stand Up Conic," by Lee Whitt, Daniel H. Wagner Associates.
"The Heat Equation on a Metal Bar with Radiating Ends," by Howard Penn, U.S. Naval Academy.
"Mathematics Without Fear," by Carol Crawford, U.S. Naval Academy.
"An Algorithm for Generating all Triangles with Integer Sides and a 60 Degree Angle," by William Hildebrand, Montgomery College.
"Some Global Metric Theorems," by Clifford Maloney, Bethesda, Maryland.
"Free Convection Phenomena," by Richard Barbieri, NASA/Goddard Space Flight Center.
"A Theorem on Basic Commutators," by Anthony Gaglioni, U.S. Naval Academy.

The new M.A.A. series of videotapes, "Mathematics at Work in Society," was screened both Friday and Saturday.

$$\leqslant 1/2^q - 2/4^q + (1/2^q)\left\{\sum_{n=q}^{\infty} m(A_n)\right\} \qquad \text{(by Note 4)}$$

$$< 1/2^q - 2/4^q + (1/2^q)\left\{\sum_{n=q}^{\infty} 1/2^n\right\} = 1/2^q = m(B) \qquad \text{(by Note 1).}$$

Therefore we have $0 < m(A \cap J)/m(J) < 1$ for all nonempty open subintervals J of I, and hence A is an approximate comb.

References

1. E. Hewitt and K. R. Stromberg, Real and Abstract Analysis, Springer-Verlag, 1965.
2. ______, Some examples of nonmeasurable sets, J. Austral. Math. Soc., 18 (1974) 236–238.
3. H. L. Royden, Real Analysis, Macmillan, New York, 1968.
4. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1974.

Who is the wise guru, and who are his disciples?

(Answer on page 129.)

$$\binom{2m+1}{m+1} = \binom{2m+1}{m}$$

and both occur in the expansion of $(1+1)^{2m+1}$. The two inequalities above imply that

$$d_n \leqslant 4^n, \qquad \forall n \in \mathbb{N},$$

as required.

Note that we may now, by familiar methods, deduce the upper bound for $\pi(N)$. From (13) we have, for any t with $1 < t \leqslant n$,

$$\prod_{t<p\leqslant n} p \leqslant d_n \leqslant 4^n$$

so that

$$(\log t)(\pi(n) - \pi(t)) \leqslant \sum_{t<p\leqslant n} \log p \leqslant n \log 4.$$

Hence

$$\pi(n) \leqslant \frac{n \log 4}{\log t} + t.$$

Put $t = n/\log^2 n$ to deduce that

$$\pi(n) \leqslant \frac{n \log 4}{\log n} + O\left(\frac{n \log\log n}{\log^2 n}\right).$$

This implies the upper bound in (2).

In conclusion, it is perhaps appropriate to point out that Theorem 3 can also be proved by the standard methods of proof. The interest here lies essentially in the rather curious nature of this proof. It is unexpected to use (i) to prove (ii), and it certainly is strange that there is no mention of primes in the proof of Theorem 3. It also seems worthwhile to point out that there are different ways to prove the identity implied by equations (7) and (8), for example, by expressing $1/x(x+1)\cdots(x+m)$ in partial fractions or by using the difference operator.

References

1. P. L. Chebyshev, Mémoire sur les nombres premiers, Oeuvres I, 1852, pp. 51–70
2. W. J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, Mass., 1977.
3. P. Erdős, Acta Universitatis Szegediensis (Szeged, Hungary), 5 (1932) 194–198.

THE WISE GURU AND HIS DISCIPLES (PAGE 125)

The central figure is André Weil. On his right is Serge Lang, and on his left Peter Swinnerton-Dyer (on the floor), and E. C. Zeeman. The picture was taken in approximately 1955.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by June 30, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2926. *Proposed by Jerrold W. Grossman, Oakland University.*

The cells in a lower triangular array are indexed by pairs of integers (i, j), where $1 \leqslant j \leqslant i \leqslant n$. There are two natural ways in which to order the cells:

$$\text{ROW ORDER:}\quad (i,j) < (i',j') \Leftrightarrow i < i' \quad \text{or} \quad (i = i' \text{ and } j < j').$$
$$\text{COLUMN ORDER:}\quad (i,j) < (i',j') \Leftrightarrow j < j' \quad \text{or} \quad (j = j' \text{ and } i < i').$$

Let $S(n)$ be the number of cells whose rank is the same in both systems.

(a) Evaluate $S(n)$.

(b) Find the values of n for which $S(n) = 4$.

(c) Find the values of n for which $S(n)$ is odd.

E 2927*. *Proposed by Clark Kimberling, University of Evansville.*

Define sequences $\{a_n\}, \{b_n\}, \{c_n\}$ inductively as follows: $a_1 = 1$, $b_1 = 2$, $c_1 = 4$, and take

$$a_n = \text{least positive integer not among } a_1,\ldots,a_{n-1}, b_1,\ldots,b_{n-1}, c_1,\ldots,c_{n-1}$$
$$b_n = \text{least positive integer not among } a_1,\ldots,a_{n-1}, a_n, b_1,\ldots,b_{n-1}, c_1,\ldots,c_{n-1}$$
$$c_n = 2b_n + n - a_n.$$

Prove or disprove that $0 < n(1 + \sqrt{3}) - b_n < 2$ for all n.

E 2928. *Proposed by Doug Wiedemann, Institute for Defense Analyses.*

Find $\lim(\sum_1^\infty x^n/n^n)$, $x \to -\infty$.

E 2929. *Proposed by Leo J. Alex, State University College, Oneonta, NY.*

Find all solutions to the equation $1 + 3^a = 5^b + 3^c$ in integers a, b, c.

E 2930. *Proposed by the editors.*

Find the largest square that can be inscribed in some triangle of area 1.

E 2931. *Proposed by Chen-Te Yen, Chung Yuan Christian University, Chung-Li, Taiwan.*

Let n, m be fixed positive integers. Suppose G is a group such that for all x, y in G, $(xy)^n = x^n y^n$, $(xy)^{n+m} = x^{n+m}y^{n+m}$, and $(xy)^{n+2m} = x^{n+2m}y^{n+2m}$. Show that G is abelian if $m = 1$ or 2. What about $m \geqslant 3$?

SOLUTIONS OF ELEMENTARY PROBLEMS

A Recurring Sequence of Points in the Affine Plane

E 2736 [1978, 682]. *Proposed by E. Ehrhart, University of Strasbourg, France.*

Let Δ be a closed triangle and $P_0, A_0, P_1, A_1, \ldots$ an infinite sequence of points in a plane. Assume that $P_i \neq P_{i+1}$, $A_i \neq A_{i+1}$, each A_i is a vertex of Δ and the midpoint of the segment $[P_i, P_{i+1}]$, and that $[P_i, P_{i+1}] \cap \Delta = \{A_i\}$.

Prove that $P_n = P_0$ for some positive n.

Solution by the proposer. The proposition will be established if we show that the orbit of the P_i for even i is finite. (For odd i, P_1 plays the role of P_0.)

It is sufficient to consider an equilateral triangle, since every triangle can be transformed to an equilateral one by an affinity, and the affinity preserves proportionality of line segments.

The point P_0 may be taken as a node in a lattice of equilateral triangles which tile the plane, each side of every such triangle parallel to and double the side of the given triangle. Each segment $P_{2i}P_{2i+2}$ is an edge in this lattice, since the product of two consecutive half turns in points A and B is the translation by the vector $2\,\overrightarrow{AB}$.

Let ABC be the given equilateral triangle, and to fix the ideas let $\overrightarrow{P_{2k}P_{2k+2}} = 2\,\overrightarrow{AB}$. The vector $\overrightarrow{P_{2k+2}P_{2k+4}}$ is equal to either $2\,\overrightarrow{AB}$ or $2\,\overrightarrow{CB}$ or $2\,\overrightarrow{CA}$. The oriented angle $P_{2k}P_{2k+2}P_{2k+4}$ is therefore equal to $\pi/3$, $2\pi/3$, or π.

We examine the polygonal line $L = P_0P_2P_4\ldots$ Evidently, when it turns, it turns always in the same sense and always through 60° or 120°. Moreover, it cannot continue straight indefinitely since (using the symbolism of the preceding paragraph) after at most a finite number of occurrences of $2\,\overrightarrow{AB}$ in the sequence, L will cross AC.

We will show that L is bounded. Then, since there exists only a finite number of lattice points in a bounded region, L will have to close, and since no point can have two predecessors it will have to close at P_0.

Each segment $P_{2k}P_{2k+2}$ determines an open half-plane which L cannot enter. It cannot, of course, cross itself since no point can have two predecessors. If it came to the prolongation of P_0P_2 in the direction of P_0, then we would follow L backwards (the original choice of P_2 was arbitrary between two points, of course). It is trapped in a bounded region containing finitely many lattice points and yet cannot terminate or repeat itself.

Thus L must eventually turn through 360° while lying wholly in the appropriate half-plane determined by P_0P_2. It also lies in a half-plane determined by every $P_{2k}P_{2k+2}$. After four such half-planes have been determined, L is confined to a bounded region.

Alternate solution by Gustaf Gripenberg, University of Wisconsin at Madison. Let the vertices of Δ be A, B, and C. For an arbitrary point P in the plane, let us define coordinates (x, y, z) as

follows: Let x be the quotient of the distance of P to BC by twice the distance of A to BC, taken positive if A and P lie on the same side of BC and negative if they lie on opposite sides. We define y and z similarly with respect to AC and AB. Let the coordinates of P_n be (x_n, y_n, z_n).

We examine $k_n = [|x_n|] + [|y_n|] + [|z_n|]$ and show that it is bounded—in fact it takes on at most two values and they are consecutive integers. This will show that all P_n lie in a bounded region of the plane.

If A is the midpoint of P_nP_{n+1}, then

$$x_{n+1} = 1 - x_n, \quad y_{n+1} = -y_n, \quad \text{and } z_{n+1} = -z_n; \tag{1}$$

similar statements hold for the other cases. Hence $x_n = \pm x_0 + k$, $k \in Z$, etc.; so in a bounded region there exists only a finite set of possible points for the sequence P_n. It must therefore repeat; but clearly if $P_{n+k} = P_k$ then $P_n = P_0$.

By (1), $|k_{n+1} - k_n| \leq 1$. We will show that if $k_{n+1} - k_n = 1$, then $k_n - k_{n-1} = -1$, or equivalently $k_{n+1} = k_{n-1}$. Since the sequence can be read in reverse, the converse is also true. This establishes that k_n takes on at most two values, and in fact cannot assume the lower value more than once consecutively.

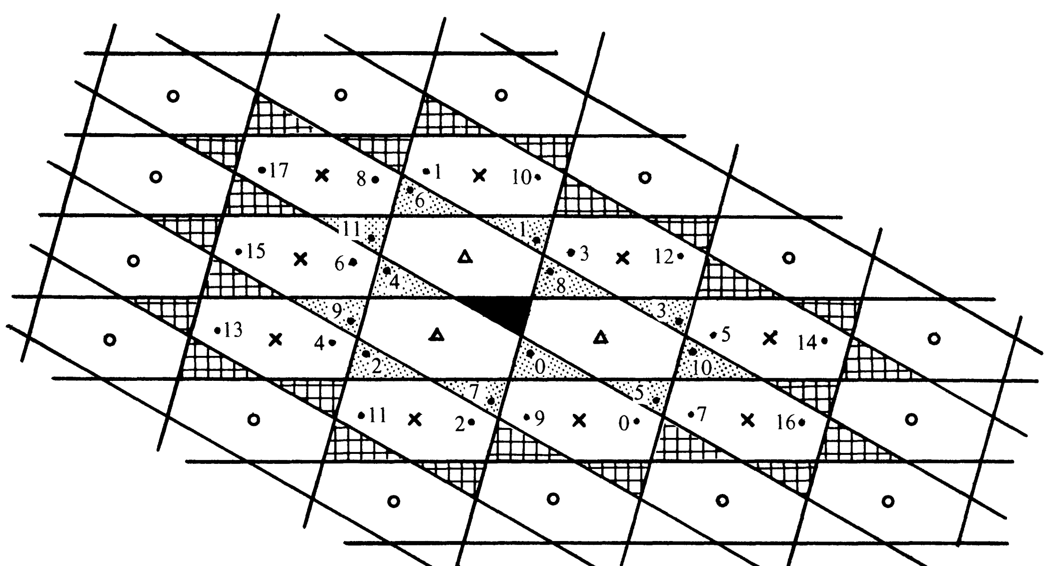
Let A be the midpoint of P_nP_{n+1}. Then from our hypothesis that $k_{n+1} = k_n + 1, [|x_{n+1}|] = [|x_n|] + 1$, and since $|x_{n+1}| = |1 - x_n|$, we infer that $x_n < 0$.

Since A is the midpoint of P_nP_{n+1}, y_n and z_n are not both positive. All three coordinates cannot be of the same sign; so one of y_n and z_n is positive, the other negative. Say $y_{n-1} < 0$ and $z_n > 0$. But this implies that B is the midpoint of $P_{n-1}P_n$ and that $y_{n-1} = 1 - y_n$; and so $[|y_{n+1}|] = 1 + [|y_n|]$. As $|z_{n-1}| = |z_n| = |z_{n+1}|$, $|x_{n-1}| = |x_n|$ and $|y_n| = |y_{n+1}|$, we get $k_{n-1} = k_{n+1}$, which was to be proved. □

Peter Addor, student, Bern, Switzerland, writes, "This problem arises from the question of stability in a three-body system. Its solution is mentioned in *The Mathematical Intelligencer*, vol. 1, #2, page 66, Remark 2. Instead of an equilateral triangle, you can take an arbitrary triangle."

Many incorrect solutions were received, and most tried to show that $P_6 = P_0$. Actually the smallest n for which $P_n = P_0$ depends heavily on the location of P_0 relative to Δ. However, n must be a multiple of 3, and it can be any multiple of 3.

In the figure below we have generated a lattice of triangles congruent to the original (which is shown in black) by successive half-turns around the vertices. P_0 cannot lie on any of the lines, as it would be impossible to prolong the sequence of P_n's to infinity and still meet the required disjointness condition.



Let us define for each region of this lattice a *shell index*. The original triangle has a shell index of 0. The three hexagons bordering it have shell index 1. The twelve triangles bordering these hexagons have shell index 2, etc. The nth shell consists of

(1) $3n$ hexagons, if n is odd;

(2) $6n$ triangles, if n is even.

If P_0 is any point in shell n, then its orbit intersects each region in the nth shell and no others. If n is even, this orbit has exactly one point in each region; hence it has $6n$ points, i.e., $P_{6n} = P_0$. Thus for the stippled triangles (shell 2), $P_{12} = P_0$; for the hatched triangles (shell 4), $P_{24} = P_0$; etc.

If n is odd, there is a special case: if P_0 is the center point of a hexagon (each of which is symmetric under a half-turn about its center), then its orbit has again exactly one point in each region of the shell—namely, the midpoints—so $P_{3n} = P_0$. Thus for the points denoted $\triangle$ in the figure (shell 1), $P_3 = P_0$; for those denoted x (shell 3), $P_9 = P_0$; for those denoted $\circ$ (shell 5), $P_{15} = P_0$; etc.

If P_0 is any other point in a hexagon, then its orbit includes two symmetrically located points in each hexagon. For such points, $P_{6n} = P_0$. Thus for points in the innermost three hexagons (excluding the $\triangle$'s), $P_6 = P_0$. For points in the next ring of nine hexagons (excluding the $\times$'s), $P_{18} = P_0$. For points in the 15 hexagons in the outermost ring shown (excluding the $\circ$'s), $P_{30} = P_0$.

The region in which P_n lies and the region next to the region in which P_{n+1} lies are always opposite (with respect to Δ). Thus the sequence seesaws around the shell. The P_n with odd indices make a continuous circuit around the shell, moving always to contiguous regions, and so do those with even indices. Typical sequences are illustrated in shells 2 and 3.

Rank 1 Matrices $A, B, A + B$

E 2851 [1981, 672]. *Proposed by Peter Ungar, New York University.*

Suppose all three matrices, A, B, $A + B$ have rank 1. Prove that either all the rows of A and B are multiples of one and the same row vector v or else all the columns of A and B are multiples of one and the same column vector w.

Solution by Noel Glick, student, Brooklyn College, New York. Since A has rank 1, clearly all the rows are multiples of one row vector $v \neq 0$. A similar remark applies to B with corresponding row vector w. Thus, if $A_i(B_i)$ denotes the ith row of $A(B)$ we have $A_1 = c_1v$, $A_2 = c_2v$, $A_3 = c_3v, \ldots,$ and $B_1 = d_1w$, $B_2 = d_2w$, $B_3 = d_3w, \ldots (c_j$ and d_j are constants). Assume that w is not a multiple of v. It follows that the column vector $c = (c_1, c_2, c_3, \ldots)$ is a multiple of the column vector $d = (d_1, d_2, d_3, \ldots)$, otherwise $A + B$ would have at least 2 independent rows. But this means that the columns of A, B, and $A + B$ are all multiples of the column vector c.

Duarte mentions the generalization (Marcus, *Finite-dimensional Multilinear Algebra*, Dekker, 1973): Any nonzero $m \times n$ matrix A over a field can be written as a sum of r rank one matrices:

$$A = \sum_{t=1}^{r} [c_{ti} d_{ti}].$$

Moreover, A has rank r if and only if each of the two sets of vectors $\{(c_{t1}, \ldots, c_{tm})\}$, $\{(d_{t1}, \ldots, d_{tn})^*\}$, $t = 1, \ldots, r$ is a linearly independent set.

Also solved by A. L. Duarte (Portugal), M. Fields (student), G. Gagola, F. Gerrish (United Kingdom), V. Hernandez (Spain), J. T. Holmes, M. Josephy (Costa Rica), M. F. Kruelle (student), J. R. Kuttler, A. Lavin (student), O. P. Lossers, Jr. (Netherlands), M. Marcus, N. Passell, J. F. Queiró (Portugual), S. Ricci, G. S. Rogers, H. Schwerdtfeger (Canada), J. Suck (Germany), W. V. Webb, H. Wolkowicz (Canada), Y.-L. Wong (Hong Kong), P.-Y. Wu (China).

ANSWERS TO "PHOTOS" ON PAGE 107

Top left: A. S. Besicovitch; top right: D. H. Lehmer; bottom left: N. E. Steenrod; bottom right: J. W. Tukey.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by June 30, 1982. The solver's full post-office address should be on each sheet.

6377. *Proposed by Kenneth R. Kellum, California State University, Sacramento.*

Suppose G is a subset of the Euclidean plane such that G meets each vertical line in exactly two points and G meets each nonvertical line in a dense set of points. Must G have a subset H such that H meets each vertical line in one point and each nonvertical line in a dense set of points?

6378. *Proposed by Yi Hong and Jingcheng Tong, Wayne State University.*

Let $f(x)$ be a real function defined on $I = [0, 1]$. Prove there are two Lebesgue measurable functions $g(x)$ and $h(x)$ such that $f(x) = g(h(x))$.

6379. *Proposed by Robert Curry and James O. Friel, California State University, Fullerton.*

Find $\liminf_{n\to\infty} |\sin n|^{1/n}$.

SOLUTIONS OF ADVANCED PROBLEMS

Location of a Zero of a Complex Polynomial

6191* [1978, 54]. *Proposed by Harry D. Ruderman, Hunter College Campus School.*

Let $P(z)$ be a monic polynomial with complex coefficients, in the complex variable z. Let $P(z_1)$ and $P(z_2)$ be in opposite quadrants I and III or II and IV. Let $z_3 = (z_1 + z_2)/2$. What is an upper bound (least, if possible) on r which will guarantee that a zero of $P(z)$ will be within a distance r from z_3?

Solution by M. Marden, University of Wisconsin–Milwaukee. We may assume, without loss of generality, that $0 < \arg P(z_1) < \pi/2$. Then $0 < \arg[-P(z_2)] < \pi/2$. If now we apply Theorem (24, 1) on page 110 of M. Marden, *Geometry of Polynomials*, with $\alpha_1 = -P(z_2)$, $\alpha_2 = P(z_1)$, $\sigma = 0$, $\gamma = \pi/2$ and K the line segment from z_1 to z_2, we conclude that at least one zero of $P(z)$ lies in the region $S(K, \pi/2n)$, bounded by two symmetric circular arcs and consisting of all points from which K subtends an angle of at least $\pi/2n$, where n is the degree of $P(z)$. Thus, on enclosing $S(K, \pi/2n)$ in the smallest circle with center at z_3, we infer that

$$0 < r \leqslant (1/2)\,|z_1 - z_2|\cot(\pi/4n).$$

However, I do not know if this is the least upper bound.

Theorem (24, 1) in fact yields the more general result that, if both points $w_1 = P(z_1)$ and $w_2 = -P(z_2)$ lie in the sector

$$\mu \leqslant \arg z \leqslant \mu + \gamma < \mu + \pi,$$

then at least one zero of $P(z)$ lies in the region $S(K, \omega)$ consisting of all points z from which segment K subtends an angle of at least $\omega = (\pi - \gamma)/n$. That is, at least one zero of $P(z)$ lies in the circle center at z_3 and radius $(1/2)\,|z_1 - z_2|\cot(\omega/2)$.

Images of Monotone Functions

6218 [1978, 500]. *Proposed by M. J. Pelling, Balliol College, Oxford, England.*

Let S be a subset of the real line R having cardinality of the continuum. Is there always a monotonic $f\colon R \to R$ such that $m^*f(S) > 0$ where m^* is outer Lebesgue measure?

Partial solution by Fred Galvin, University of Kansas. There are counterexamples under the

assumption that R is not the union of fewer than $2^{\aleph_0}$ nowhere dense sets. This assumption, called the "Strong Baire Category Theorem" or SBCT [D. A. Martin and R. M. Solovay, *Internal Cohen extensions*, Ann. Math. Logic, 2 (1970) 143–178], is weaker than the continuum hypothesis or Martin's axiom, but is not provable in ZFC.

Martin and Solovay noted (p. 177) that, assuming SBCT, there is a set $S \subseteq R$ of cardinality $2^{\aleph_0}$ such that $f(S)$ has Lebesgue measure 0 for every continuous function $f: R \to R$, a result obtained earlier by W. Sierpiński assuming the continuum hypothesis. The example can easily be improved so that $f(S)$ has Lebesgue measure 0 for every function $f: R \to R$ having only a meager set of discontinuities, in particular, for every monotonic function. In order to give the counterexample in a sharper form, I need the following definitions.

A set $S \subseteq R$ is a *generalized Lusin set* if $|S \cap N| < 2^{\aleph_0}$ for every nowhere dense set $N \subseteq R$. (It follows that $|S \cap M| < 2^{\aleph_0}$ for every meager set $M \subseteq R$.)

A set $S \subseteq R$ is a *Rothberger set* if, for every sequence $\mathfrak{U}_1, \mathfrak{U}_2, \ldots, \mathfrak{U}_n, \ldots$ of open covers of S, it is possible to choose $U_1 \in \mathfrak{U}_1, U_2 \in \mathfrak{U}_2, \ldots, U_n \in \mathfrak{U}_n, \ldots$ so that $S \subseteq \cup_{n=1}^{\infty} U_n$. Rothberger sets are also called sets with the property C'' [Fritz Rothberger, *Sur les familles indénombrables de suites de nombres naturels et les problèmes concernant la propriété C*, Proc. Cambridge Philos. Soc., 37 (1941) 109–126]. Clearly, a Rothberger set has Lebesgue measure 0.

The Rothberger property is not necessarily hereditary; see Rothberger's Corollaire 2(2). I will call $S \subseteq R$ a *hereditary Rothberger set* if every subset of S is a Rothberger set.

LEMMA. *Assuming* SBCT, *every subset of R having cardinality* $< 2^{\aleph_0}$ *is a Rothberger set.*

This Lemma is essentially due to Rothberger; it is obtained by replacing the cardinal $\aleph_1$ in his Lemma 8 by an arbitrary cardinal $< 2^{\aleph_0}$. For a concise proof of the Lemma, see Theorem 1.4 of Richard Laver, *On the consistency of Borel's conjecture*, Acta Math., 137 (1976) 151–169, where however SBCT is stated in the needlessly strong form that the union of $< 2^{\aleph_0}$ meager sets is meager.

For a function $f: R \to R$, let $C(f) = \{x \in R : f \text{ is continuous at } x\}$.

THEOREM. *Assume* SBCT. *Then*:

(1) *there is a generalized Lusin set of cardinality* $2^{\aleph_0}$;

(2) *if S is a generalized Lusin set, and if* $f: R \to R$ *is such that* $C(f)$ *is dense in R, then* $f(S)$ *is a hereditary Rothberger set.*

Part (1) is trivial; see p. 174 of Martin and Solovay. For (2) it is enough to prove that $f(S)$ is a Rothberger set, since every subset of a generalized Lusin set is a generalized Lusin set. Consider any sequence $\mathfrak{U}_1, \mathfrak{U}_2, \ldots, \mathfrak{U}_n, \ldots$ of open covers of $f(S)$. Let D be a countable dense subset of $S \cap C(f)$. Choose $U_{2n} \in \mathfrak{U}_{2n}$ $(n = 1,2,3,\ldots)$ so that $f(D) \subseteq U = \cup_{n=1}^{\infty} U_{2n}$. Since $D \subseteq C(f)$, there is an open set $V \supseteq D$ such that $f(V) \subseteq \bar{U}$. Now $S - V \subseteq [S \cap (\bar{V} - V)] \cup [S - C(f)]$. Since S is a generalized Lusin set, and since $\bar{V} - V$ is nowhere dense and $C(f)$ is a dense G_δ, it follows that $|S - V| < 2^{\aleph_0}$. Since $f(S) - U \subseteq f(S - V)$, we have $|f(S) - U| < 2^{\aleph_0}$. By the Lemma, $f(S) - U$ is a Rothberger set. Thus we can choose $U_{2n-1} \in \mathfrak{U}_{2n-1}$ $(n = 1,2,3,\ldots)$ so that $f(S) - U \subseteq \cup_{n=1}^{\infty} U_{2n-1}$, i.e., $f(S) \subseteq \cup_{n=1}^{\infty} U_n$.

William J. Gorman III (Purdue University) and John C. Morgan II (California State Polytechnic University, Pomona) note that a negative solution, under the assumption of the Continuum Hypothesis, was given by W. Sierpiński, *Hypothèse du Continu*, 2nd ed., 1956, p. 49. Since the Strong Baire Category Theorem can hold when the Continuum Hypothesis fails, the solution by Galvin gives additional information.

Branko Curgus (University of Sarajevo, Yugoslavia) gives essentially the argument of Sierpiński, assuming the existence of an uncountable subset S of R concentrated on a countable set (see C. A. Rogers, Hausdorff Measures, Cambridge University Press, 1970, p. 74), an assumption independent of ZFC.

Structure of Finite Rings.

6284 [1979, 869]. *Proposed by William P. Wardlaw, U.S. Naval Academy.*

Let R be a finite ring with more than one element and with no nonzero nilpotent element.

Show that R is a direct sum of fields.

(This generalizes Wedderburn's theorem that a finite division ring is a field.)

Composite solution. Since R is finite with zero radical, the Wedderburn Theorem implies that R is a finite direct sum of complete matrix rings over finite division rings, that is, fields. Since R has no nonzero nilpotent elements, each of these matrix rings is simply a finite field.

Solutions from first principles, proofs using a variety of other known theorems, as well as proof of known generalizations, were given.

Solved by Daniel D. Anderson, Howard E. Bell, D. M. Bloom, F. S. Cater, Chico Problem Group, Charles Chouteau, J. A. Cuenca (Spain), Enzo R. Gentile (Argentina), Anthony V. Geramita, Robert Gilmer, Bruce Glastad, Ronald Gutschow, A. A. Jagers (Netherlands), Richard Johnsonbaugh, W. G. Leavitt, Jonathan Leech, Steve Ligh, Jiang Luh, Robert A. Melter, Robert L. Miller, W. K. Nicholson, Dalton Orr, Barbara L. Osofsky, John Petro, Douglas F. Rall, Harry F. Smith, Earl J. Taft, Rony Teitler (England), Ernst Trost (Switzerland), Chester E. Tsai, Gregory P. Wene, Edward T. Wong, Qazi Zameeruddin (India), and the proposer.

REVIEWS

EDITED BY JOHN H. EWING

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE.

Old and New Problems in Combinatorial Number Theory. By P. Erdős and R. L. Graham. L'Enseignement Mathématique Mon. No. 28, Geneva, 1980. 128 pp., 38SF.

HEINI HALBERSTAM

Department of Mathematics, University of Illinois, Urbana, IL 61801

The readers of this MONTHLY are certainly no strangers to problems, and they will therefore be especially glad to welcome another collection of problems bearing the name of Paul Erdős, chief paramount of problem posers, and of R. L. Graham, a distinguished member of the group who, with Erdős, has in our own time raised the asking of questions to a fine art. Readers will marvel anew, over and over again, as I have done, how on each page, with each deft, masterful flick of the mathematical kaleidoscope, some new speculation about numbers is born to challenge our understanding.

The present collection is arranged, apart from one inevitable section of miscellanea, into seven chapters; and each chapter evolves, from some one basic result, a host of problems that probe into the numerous open ramifications of the original result. In the process there is revealed an imposing network of unsuspected connections; a part of the vast inchoate expanse of mathematical ignorance begins in this way to acquire a geography.

The second chapter (the first is an introduction) deals with Van der Waerden's theorem, according to which, if the positive integers are divided in any way into two classes, at least one class contains an arbitrarily long arithmetic progression (A.P.). Here is a sample thread of problems from the thirteen pages devoted to variations of the central theme. Let A be a set of positive integers such that $\Sigma_{a\epsilon A}1/a = \infty$. Is it true that A must contain arbitrarily long A.P.'s? The authors describe this problem as "quite deep" and Erdős offers \$3,000 for its solution. If true, this result would imply Szemerédi's theorem: Let $r_k(n)$ denote the least integer r so that, if $1 \leqslant a_1 < \cdots < a_r \leqslant n$, the sequence of a_i's contains an A.P. of k terms. Then $r_k(n) = o(n)$. The authors suggest that perhaps $r_k(n) = O(n(\log n)^{-t})$ for every t. This would imply that, for every k, there are k primes which form an A.P. The longest known A.P. of primes has 17 terms. Let $\{a + kd : 0 \leqslant k \leqslant t\}$ with $a + kt < x$ consist entirely of primes. Is it true that $t = o(\log x)$? What

can one say about t if one requires only a positive proportion of terms to be prime?

The third chapter deals with covering congruences: i.e., a (finite) system of r residue classes a_i mod n_i with $1 < n_1 < \cdots < n_r$ such that every integer satisfies at least one of the congruences $x \equiv a_i \bmod n_i$. The original example of such a system was 0 mod 2, 0 mod 3, 1 mod 4, 3 mod 8, 7 mod 12, and 23 mod 24. Here the central theme is the open question: Given any c, is there a system of covering congruences with $n_1 \geq c$? Choi has constructed a system with $n_1 = 20$. Other questions come thick and fast. Is there a system with all the n_i odd? This would be true if there existed a system with no n_i dividing any other n_j; but that is also an open question. Given n_1, let $r = f(n_1)$ be the length of the shortest system of covering congruences starting with a congruence mod n_1. If $f(n_1)$ is finite, is it true that $f(n_1)n_1^{-k} \to \infty$ for every k? Is there some growth condition on the n_i which ensures that an associated covering system exists; e.g., is it sufficient that $\Sigma_{n_i \leq x} 1/n_i$ grows fast enough? Covering systems can be used to show that there exist sequences of integers, generated by linear recurrences, none of which is prime. There are also connections with representability of integers as sums of a prime and a number of powers of 2 (the topic derives from here, in fact), and with polynomial irreducibility questions.

The fourth chapter deals with unit fractions: the representation of rationals a/b in the form $\Sigma_{i=1}^{r} 1/x_i$ with $x_1 \leq x_2 \leq \cdots \leq x_r$, a question that goes back to Egyptian arithmetic and has spawned many interesting inquiries. Just two illustrations: if $0 < x_1 < \cdots < x_n$ and $1/x_1 + \cdots + 1/x_n = 1$, it is known that $\max(x_{k+1} - x_k) > 1$, but is it true that $\max(x_{k+1} - x_k) \geq 3$? Equality can occur (viz., $1 = \frac{1}{2} + \frac{1}{3} + \frac{1}{6}$), but does it occur ever again, or infinitely often? For a fixed $c > 0$, let $S_c \subseteq \{1, 2, \ldots, n\}$ with $|S_c| \geq cn$. Is it true that there is a function $f(c)$ so that some sum $\Sigma_{s \in S_c} 1/s$ has denominator at most $f(c)$?

The fifth chapter is about bases of integers and related problems. This is a classical problem area and the authors do not give an exhaustive account. But there is much here that is less well known. For example, this old problem of Dickson: Starting from a set $A_k = \{a_1 < \cdots < a_k\}$, define a_{n+1} for $n \geq k$ to be the least integer exceeding a_n which is not of the form $a_i + a_j$ $(i, j \leq n)$. Is it true that the sequence of differences $a_{m+1} - a_m$ is eventually periodic? We are told that even starting from $\{1, 4, 9, 16, 25\}$ requires thousands of terms to reach periodicity.

The sixth chapter is about completeness of sequences: for a sequence $A = (\alpha_1, \alpha_2, \ldots)$ of real numbers, let $P(A)$ denote the set of all numbers arising as sums of distinct elements of A. A sequence A of integers is complete if $P(A)$ contains all sufficiently large integers. A sample: Let α and β be positive reals with α/β irrational and take $A = ([\alpha], [\beta], [2\alpha], [2\beta], \ldots, [2^n\alpha], [2^n\beta], \ldots)$. Is A complete? And what if 2 is replaced by γ with $1 < \gamma < 2$?

Chapter 7 focuses on irrationality and transcendence of various special series. For example, it is known that $\Sigma_n p_n/n!$ is irrational (where p_n is the nth prime), but is the same true of $\Sigma_n p_n/2^n$? The authors speak of this as probably hopeless; but they seem more optimistic of proving as much for $\Sigma_q q/2^q$ where q ranges over the square-free integers, and also $\Sigma_n v(n)/2^n$, where $v(n)$ denotes the number of prime divisors of n. Apropos of the latter,

$$\sum_n \frac{d(n)}{2^n} = \sum_n \frac{1}{2^n - 1}$$

is known to be irrational; but what about

$$\sum_n \frac{1}{2^n - 3} \quad \text{or} \quad \sum_n \frac{1}{n! - 1}?$$

Chapter 8 deals with unconventional diophantine problems. Just to give a flavor, here are some problems that are simple to state: Is it true that $\binom{2n}{n}$ is never square free if $n > 4$? Is $\left(\binom{2n}{n}, 105\right) = 1$ infinitely often? We have $\binom{21}{2} = 2 \cdot 3 \cdot 5 \cdot 7$. Can $\binom{n}{2}$ be the product of consecutive primes infinitely often? For any set A of n positive integers, is it true that max (for

$a, a' \in A)$ $a/(a, a') \geqslant n$? Let $w(n)$ denote the number of solutions in a, b, c, and d of $n = 2^a + 3^b + 2^c \cdot 3^d$; is $w(n)$ bounded?

Chapter 9 is a large selection of miscellaneous problems, Chapter 10 updates information about some of the problems in Erdős's 1963 collection (Monograph No. 6) in the same periodical, and on the last page answers or partial answers are provided to three of the questions in the present collection.

One last word: On page 10 the authors refer to a forthcoming book by Croft and Guy, *Unsolved Problems in Intuitive Mathematics*. This has been in preparation for many years and the collection is now so large that the portion on number theory will come out shortly, under Guy's name, as Volume I.

Against Infinity: An Anthology of Contemporary Mathematical Poetry: 75 compositions. Edited by Ernest Robson and Jet Wimp. Primary Press, 1979. 90 pp. Box 105A, Parker Ford, PA 19457. $17.00 (cloth), ISBN 0-934982-00-7. $8.95 (Paper), ISBN 0-934982-01-5. (Telegraphic Review, December 1979.)

MARION COHEN

Department of Mathematics, Temple University, Philadelphia, PA 19122

This title makes me think of Walt Whitman's famous and much-loved poem "When I Heard the Learn'd Astronomer", along with William Blake's "Tyger Tyger burning bright" and that Tyger's "fearful symmetry." Fragments of Babette Deutsch, Ruth Whitman, and Adrienne Rich also float through my mind, borne by currents of metaphor derived unmistakably from science. Lucretius, with his childlike "Order of the Universe," and Archimedes, with what he wrote in the sand instead of love letters, especially vibrate the heartstrings of this poet-mathematician. And cause me to indulge in a personal childhood memory: That poem called "The Parabola," author unremembered (in fact, entire poem unremembered)—a long long poem which I, then a budding mathematician but only a seed of a poet, naively, or perhaps bravely, chose to read aloud for an English assignment; I stood up and recited that pearl to the class of squirming chattering non-poet non-mathematician swine, read out loud, probably *too* loud, and experienced the first pangs, the first hint, that math and poetry are almost too closely related for comfort.

Mathematical poetry has certainly been in existence for some time (in fact, $\forall t$!). Even before I was a poet, I used to say that, to me, *all* math is poetry. And now that I've authored a collection of my own math-poetry (*The Weirdest Is the Sphere*, Seven Woods Press, N.Y., 1979), I'm inclined to hold to that statement. Although not all poetry is math, there is no lack of mathematical poetry. However, the bringing together between two covers of math-poetry from various writers is something new. And I welcome the recognition of any phenomenon, this axiom of choice which allows us to proceed from the each to the all.

Against Infinity is a significant book. The myth that logic and emotion have empty intersection must be dispelled. (I would even venture to conjecture that the intersection is dense!) I still run into fellow poets and fellow mathematicians who exclaim, "What!—You do math *and* poetry?! What an odd combination!" Good math and good poetry feed on each other. The world needs to know that. The issue is especially relevant considering the recent upsurge of Math Anxiety clinics throughout the country. (In my own workshop at Temple University, *Against Infinity*, as well as my own "Sphere," is used as course material, and with great success.)

Although the quality of writing in *Against Infinity* is uneven, there are definitely enough good pages to justify its publication. "The Corporal Who Killed Archimedes" by Miroslav Holub is an example. I like it because it's math, it's poetry, and it's political. And it's short enough to quote in full:

in one bold stroke
he massacred the circle, the
tangent, the point of
intersection at infinity

on pain of
quartering he banned
numbers
from three on up

in Syracuse he now
heads a college of
philosophers squats

on his halberd
and for another thousand years writes

one two
one two
one two
one two

Another favorite of mine is the same author's "Zito the Magician," a well-crafted parable about the nature of absolute truth (although Zito should have complex angles in his bag of tricks!). A third good poem is Jacqueline Lapidus's "Several Hypotheses and a Proposition." This, to me, is noteworthy for its feminism and for its imagery, which relies truly on math, and not mysticism. Still other penetrating work is by Peter Meinke, Linda Pastan, Naomi Replansky, Zahrad, Lillian Morrison, and the editors, Ernest Robson and Jet Wimp. (Jet Wimp is a mathematician.)

Neither the editors nor I claim to relate to everything in *Against Infinity*. In particular, I, not having tried, nor wanted to try, my hand at visual poetry (yet?), cannot warm up to it. To me it seems that math itself does a far better job. However, I do see the attraction in, e.g., what appears on pages 17, 21, and 67.

Let me repeat: This is a significant book. It can be read on many levels, and applied in a multitude of classroom situations, among them the above-mentioned Math Anxiety workshops, high-school courses, and college math-survey courses. It has, in fact, already been used in all of these. Both poetically and mathematically, much of the content is on a college level, so the book is quite accessible to many people.

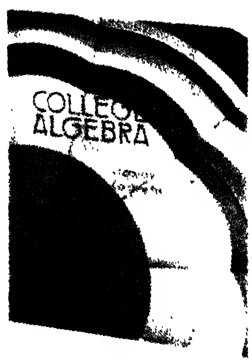
ACKNOWLEDGMENT

The editors thank the following individuals who have refereed manuscripts between mid-1980 and mid-1981: O. G. Aberth, J. F. Adams, R. A. Adams, R. A. Adler, R. L. Adler, P. Ahern, C. D. Ahlbrandt, M. O. Albertson, H. L. Alder, B. R. Alspach, D. D. Anderson, R. D. Anderson, T. W. Anderson, G. E. Andrews, N. C. Ankeny, R. F. Arens, M. Artin, R. B. Ash, R. A. Askey, C. E. Aull, C. W. Ayoub, R. G. Ayoub, D. W. Bailey, R. Balasubramanian, S. B. Bank, M. S. Baouendi, P. T. Bateman, G. Baumslag, R. A. Beaumont, E. F. Beckenbach, P. R. Beesack, L. W. Beineke, E. A. Bender, G. M. Bergman, E. R. Berkson, B. C. Berndt, S. J. Bezuszka, G. Birkhoff, D. Blackwell, B. E. Blank, T. K. Boehme, E. D. Bolker, F. F. Bonsall, W. W. Boone, W. M. Boothby, C. J. R. Borges, W. E. Boyce, L. R. Bragg, M. D. Bramson, F. G. Brauer, G. U. Brauer, G. E. Bredon, J. E. Brennan, J. P. Breezin, J. D. Brillhart, D. R. Brown, E. Brown, M. R. Brown,

W. D. Brownawell, P. J. Browne, A. M. Bruckner, R. C. Buck, R. T. Bumby, R. Bumcrot, D. M. Burton, E. Calabi, J. W. Cannon, G. T. Cargo, L. Carlitz, B. C. Carlson, T. J. Carlson, J. W. S. Cassels, R. A. Chaffer, G. Chartrand, L. Chastkofsky, P. R. Chernoff, C. C. Chicone, T. S. Chihara, C. O. Christenson, H. Cohen, P. Colwell, S. D. Comer, W. W. Comfort, E. H. Connell, L. Cooper, L. J. Corwin, T. M. Cover, C. C. Cowen, J. R. Cowles, M. J. Cowles, J. T. Cox, H. S. M. Coxeter, J. Cronin, C. W. Curtis, D. W. Curtis, A. Cusmariu, R. S. Dahiya, A. E. Danese, R. B. Darst, J. Dauns, J. W. Dawson, A. Devinatz, H. G. Diamond, R. P. Dilworth, P. G. Dixon, J. D. Dollard, W. F. Donoghue, R. S. Doran, L. L. Dornhoff, J. Douglas, R. D. Driver, D. S. Drucker, G. F. D. Duff, J. Dugundji, R. T. Durrett, P. F. Duval, M. Dwass, P. Dwinger, J. Eells, J. A. Eidswick, P. D. T. A. Elliot, R. S. Ellis, E. E. Enochs, R. J. Evans, L. Evens, J. L. Fernândez, A. M. Fink, R. L. Finney, D. T. Flesner, T. V. Fossum, D. J. Foulis, J. S. Frame, S. Friedland, A. Friedman, L. Fuchs, S. A. Gaal, J. Galambos, F. Galvin, T. W. Gamelin, K. M. Garg, S. J. Garland, G. Gasper, B. L. Gelbaum, H. Gerber, L. E. Gerber, J. L. Gersting, P. M. Gibson, L. Gillman, I. L. Glicksberg, M. Goldberg, R. R. Goldberg, L. J. Goldstein, S. W. Golomb, N. D. Goodman, W. B. Gordon, H. W. Gould, C. C. Graham, R. L. Graham, R. N. Grassl, M. J. Greenberg, H. B. Griffiths, G. G. Grinstein, B. Grofman, F. I. Gross, K. I. Gross, E. Grosswald, B. Grunbaum, S. P. Gudder, H. W. Guggenheimer, V. W. Guillemin, A. P. Guinand, M. Guzman, S. Haber, C. R. Hadlock, M. G. Hahn, J. E. Hall, M. Hall, J. M. Hammersley, F. Harary, C. M. Harris, P. Hartman, L. Hay, M. L. Hemler, J. M. Henle, J. P. Hennefeld, E. Hewitt, D. A. Higgs, A. M. Hobbs, J. G. Hocking, R. E. Hodel, R. V. Hogg, R. A. Honsberger, F. M. Hoppe, K. Hrbacek, L. Janos, T. J. Jech, M. A. Jodeit, M. J. Johnson, N. L. Johnson, P. E. Johnson, R. F. Johnsonbaugh, F. B. Jones, P. S. Jones, P. C. Kainen, J. H. Kalbfleisch, W. Kaplan, S. Karlin, H. E. Kasube, J. L. Kelley, R. B. Kellogg, P. Kiss, V. L. Klee, W. J. Knight, D. E. Knuth, N. I. Koblitz, W. K. Kohnen, I. I. Kolodner, J. D. E. Konhauser, R. D. Kopperman, R. R. Korfhage, H. L. Krall, E. O. Kreyszig, R. J. Kryscio, K. Kunen, A. E. Labarre, G. J. Lallement, J. A. Lavita, D. H. Lehmer, E. Lehmer, G. Letac, D. J. Lewis, A. E. Livingston, J. D. Logan, A. J. Lohwater, J. Q. Longyear, G. Lorentz, R. Lovett, W. F. Lucas, D. G. Malm, W. R. Mann, M. Marden, A. W. Marshall, G. E. Martin, J. M. Masley, S. B. Maurer, N. H. McCoy, P. A. McCoy, R. D. Meyer, G. Miel, E. C. Milner, T. K. Milnor, S. P. Mohanty, R. A. Mollin, H. L. Montgomery, M. S. Montgomery, C. S. Morawetz, W. O. J. Moser, M. F. Moss, B. Muckenhoupt, D. H. Mugler, M. L. Murrill, R. V. Nagisetty, D. J. Newman, J. Neveu, P. E. Ney, L. Nirenberg, I. Niven, J. K. Oddson, J. M. H. Olmsted, R. Osserman, J. C. Oxtoby, G. D. Parker, E. Parzen, N. T. Peck, R. S. Phillips, M. A. Pinsky, C. Pomerance, C. Pommerenke, J.R. Porter, T. E. Price, W. M. Priestley, M. H. Protter, E. T. Quinto, R. A. Raimi, L. B. Rall, D. Redmond, S. Reich, B. E. Rhoades, P. Ribenboim, L. F. Richardson, J. Riordan, T. J. Rivlin, D. A. Robinson, V. K. Rohatgi, S. M. Roman, A. Rosa, N. J. Rose, M. Rosenfeld, L. A. Rubel, W. Rudin, D. G. Saari, A. J. Sachs, D. H. Saracino, A. Sathaye, U. V. Satyanarayana, M. M. Schacher, J. F. Schell, A. H. Schoen, I. J. Schoenberg, A. H. Schoenfeld, H. J. Schultz, S. Schuster, A. Schwartz, A. Seidenberg, J. L. Selfridge, S. W. Semmes, D. B. Shaffer, S. M. Shah, L. W. Shapiro, J. K. Shaw, M. C. Sholander, D. B. Singmaster, D. R. Smart, M. K. Smith, P. W. Smith, E. Snapper, A. E. Solow, L. R. Sons, J. Spanier, E. L. Spitznagel, R. P. Stanley, C. M. Stanton, D. W. Stanton, N. K. Stanton, M. J. Steele, S. Sternberg, A. Sterrett, K. B. Stolarsky, A. H. Stone, G. Strang, B. Subramanian, E. R. Swart, R. H. Szczarba, E. J. Taft, O. Taussky-Todd, S. K. Thomason, M. O. Thompson, L. Tornheim, C. Tretkoff, M. D. Tretkoff, S. Y. Trimble, C. P. Tsokos, A. C. Tucker, W. T. Tutte, E. C. Turner, S. V. Ullom, W. R. Utz, J. D. Vaaler, C. Vander Eynden, E. van Douwen, J. van Yzeren, E. A. Walker, L. E. Ward, R. H. Warren, M. S. Waterman, W. E. Watkins, B. M. Weinstock. A. J. Weir, A. Weitsman, K. W. Weston, W. T. Whitley, H. Widom, S. M. Wiegand, A. Wilansky, R. L. Wilder, H. S. Wilf, S. Willard, H. C. Wiser, E. S. Wolk, D. J. Wright, J. A. Yorke, T. Zaslavsky, D. Zeitlin, K. Zeller, I. B. Zibman.

D.C. Heath has the Answer

integration by substitution, aid student understanding. More than 5,000 problems provide abundant opportunity for skill development and reinforcement. Gillett is unsurpassed in thoroughness, clarity, and effectiveness in the classroom.

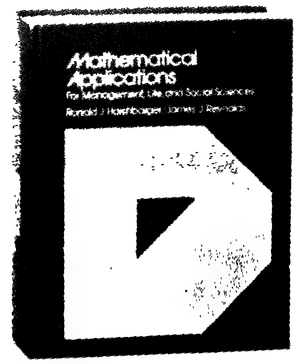


NEW! Available January 1982

College Algebra

Charles Henry Brase, *Regis College*
Corrinne Pellillo Brase,
Arapahoe Community College
1982 Cloth 480 pages
Instructor's Guide

Designed for the variety of backgrounds, interests, and mathematical abilities of today's algebra students, this text features
- a graduated problem-solving approach with more than 2,400 problems,
- interesting, practical applications in examples, exercises, and review problems,
- answers for all odd-numbered problems at the back of the text, and
- a review of basic algebra and a diagnostic pre-test.



Mathematical Applications for Management, Life, and Social Sciences

Ronald J. Harshbarger and James J. Reynolds
both of *The Pennsylvania State University*
1981 Cloth 608 pages
Selected Solutions Manual

Providing solid coverage of mathematical skills and concepts, this book features more than 800 relevant applications in management, life, and social sciences. The 2,000 problems give students ample opportunity to practice skills in practical settings related to their intended careers. An index of selected applications, chapter warm-ups testing prerequisite skills, chapter objectives and the step-by-step organization make this an effective, student-oriented presentation of mathematics.

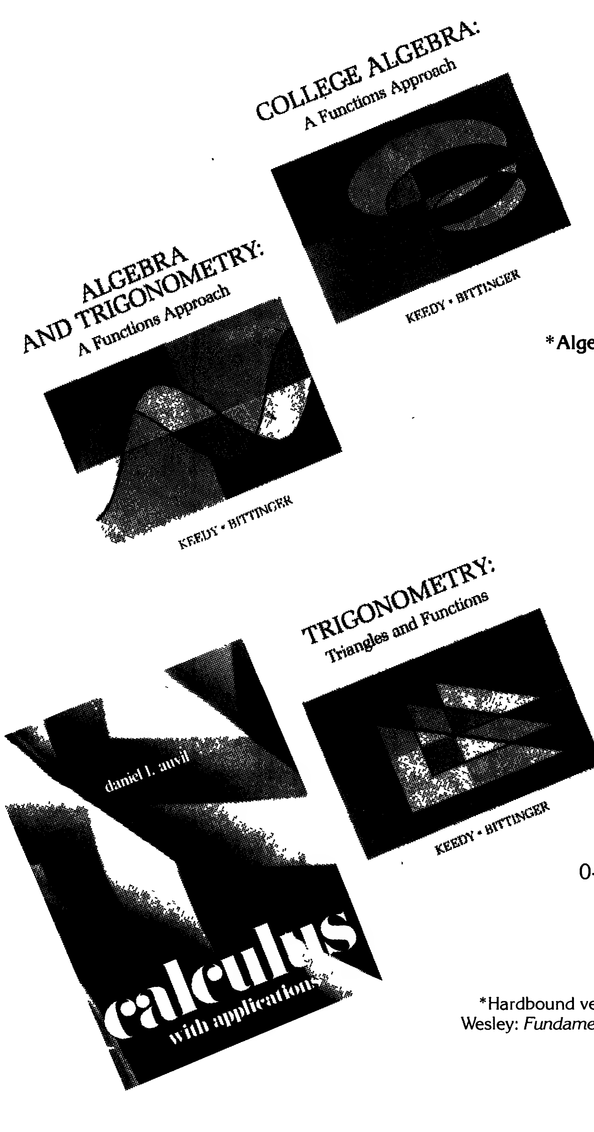
For details or sample copies,
call us toll free: 800-225-1388
In Massachusetts, call collect:
617-862-6650, ext. 1344

A Raytheon Company

HEATH

D.C. HEATH AND COMPANY
College Division
125 Spring Street
Lexington, MA 02173

Take your pick.



Precalculus and Calculus Texts

***College Algebra: A Functions Approach, Third Edition**
M. L. Keedy, Purdue University
Marvin L. Bittinger, Indiana University-Purdue University at Indianapolis
0-201-13400-4, Paper, 512 pp. (tent.), available February 1982
Student Solutions Manual, Test Booklet, and Answer Booklet available.

***Algebra and Trigonometry: A Functions Approach, Third Edition**
M. L. Keedy, Purdue University
Marvin L. Bittinger, Indiana University-Purdue University at Indianapolis
0-201-13404-7, Paper, 640 pp. (tent.), available January 1982
Student Solutions Manual, Test Booklet, and Answer Booklet available.

Trigonometry: Triangles and Functions, Third Edition
M. L. Keedy, Purdue University
Marvin L. Bittinger, Indiana University-Purdue University at Indianapolis
0-201-13408-x, Paper, 288 pp. (tent.), available February 1982
Student Solutions Manual, Test Booklet, and Answer Booklet available.

Calculus With Applications
Daniel L. Auvil, Kent State University, Stark Campus
0-201-10063-0, Hard, 448 pp. (tent.), available February 1982
Student Supplement available.

*Hardbound versions of these texts are also available from Ac 'ison-Wesley: *Fundamental College Algebra, Second Edition* and *Fundamental Algebra and Trigonometry, Second Edition*

HOUGHTON MIFFLIN UPDATE:

Teaching Elementary School Mathematics: An Active Learning Approach

Harold H. Lerch
University of Illinois, Urbana
398 pages • cloth • Activities Handbook (see below) • Instructor's Manual • 1981

Using an active learning approach, Lerch treats both methods and general issues of teaching elementary school mathematics—the scope of programs, analysis of programs, selection and evaluation of teaching procedures, and more.

He develops each content area by themes, performance objectives, and teaching strategies.

Lerch's coordinated activities handbook (see below) is the perfect partner for his text.

Active Learning Experiences for Teaching Elementary School Mathematics

Harold H. Lerch
461 pages • paper • 1981

For pre-service and in-service teachers, Lerch offers nearly 700 activities, projects, and games to use in an active learning approach to teaching elementary school mathematics.

Lerch provides a statement of purpose, a list of materials needed, and procedures for each activity.

An ideal companion to Lerch's textbook (see above), the handbook can be used with other mathematics texts.

Teaching Mathematics: Methods and Content Second Edition

Fredricka K. Reisman
University of Georgia
512 pages • cloth • Instructor's Manual 1981

Concentrating on actual classroom practice, Reisman emphasizes fundamental mathematical processes rather than elements of the "New Math." About 480 mini-lesson plans and special activities expose pre-teachers to proven instructional techniques for elementary and middle school.

The extensively revised Second Edition features new and expanded treatment of topics such as planning for instruction, the mathematics curriculum, classroom teaching, and exceptional children, as well as assessment and diagnosis. New chapters on negative numbers and word problems have also been added.

Form and Style: Theses, Reports, Term Papers Sixth Edition

William Giles Campbell
Stephen Vaughan Ballou
Carole Slade, Columbia University
About 208 pages • paper • spiralbound
Just published

A thoroughly revised edition of the highly successful, detailed guide to all aspects of developing, writing, and typing papers. Reorganization, expanded discussions, and added examples and facsimile pages make the Sixth Edition an even more efficient resource, for both graduate and undergraduate papers.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 3 March 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

HERMANN GRASSMANN AND THE PREHISTORY OF UNIVERSAL ALGEBRA

DESMOND FEARNLEY-SANDER

Mathematics Department, University of Tasmania, GPO Box 252C, Hobart 7001, Tasmania, Australia

I should begin by apologizing for considering a subject, Universal Algebra, of which I know little. I was brought to Universal Algebra against my will, as it were, by Hermann Grassmann, and the main point of this paper is to describe a piece of Grassmann's work and to ask those who know the subject better than I do whether it may be said to anticipate Universal Algebra.

Before doing that, though, I would like to set the scene by first briefly describing a typical piece of Universal Algebra, and then giving a rapid sketch of the prehistory of Universal Algebra—by which I mean simply those developments that predate the definitive formulation of the subject but which, with hindsight, may be seen to have anticipated or influenced it.

1. Universal Algebra. I am not going to assume familiarity with Universal Algebra—the few remarks that I will make now should suffice to make what follows intelligible. To save argument I will define the subject as comprising what is to be found in modern books bearing the title *Universal Algebra*, such as the ones by Cohn [**5**] and Grätzer [**8**].

Let me outline some terminology and a typical construction.

A *universal algebra* is a set G together with a system of n-ary operations for G; here n may vary and the number of operations may be infinite. (An n-ary operation is simply a function $G^n \to G$.)

These are the objects of study of the subject, Universal Algebra. They include, for example, groups, rings, linear spaces and lattices (and, with slight modification of the definition, also fields, projective planes, etc.).

In the case where there is just a single binary operation (denoted here by juxtaposition) we have a *groupoid*. Given a set of symbols $S = \{x_1, x_2, \ldots, x_n\}$, the set

$$G = \{x_1, x_2, \ldots, x_n, (x_1x_1), (x_1x_2), \ldots, (x_nx_n), (x_1(x_1x_1)), (x_1(x_1x_2)), \ldots, \\ (x_1(x_nx_n)), (x_2(x_1x_1)), \ldots, (x_n(x_nx_n)), ((x_1x_1)x_1), ((x_1x_1)x_2), \ldots\}$$

obtained by repeated juxtaposition of the symbols already written down, forms a groupoid in a natural way (the binary operation being juxtaposition). This is the *free groupoid* on S.

Now let A be the set of all (formal, finite) linear combinations of elements of G (with coefficients from a field R):

$$A = \{\Sigma\alpha_j X_j \colon \alpha_j \in R, X_j \in G\}.$$

A becomes a linear space if we define addition and multiplication by scalars in the natural way, and indeed it becomes a (nonassociative) linear algebra over R if we define

$$(\Sigma\alpha_i X_i)(\Sigma\beta_j X_j) = \Sigma\alpha_i\beta_j(X_iX_j),$$

where we need an obvious convention to cope with the fact that the indices i and j may range over different finite sets of natural numbers. This is the *free linear algebra* on S over R.

The free algebra A may be put to work in the following way. A linear algebra B is said to satisfy the *law* (or *universal relation*)

$$\Sigma\alpha_j X_j = 0,$$

Desmond Fearnley-Sander is a Senior Lecturer in Mathematics at the University of Tasmania. He previously held positions at California State College at San Bernardino and at the University of Western Australia. His main mathematical interests are linear algebra and its history, and the development and application of the approach to geometry originated by Hermann Grassmann. With J. P. O Silberstein of the University of Western Australia, he is currently working on a translation of Grassmann's *Ausdehnungslehre* of 1862.—(*Editors*)

where $\Sigma\alpha_j X_j \in A$, if a valid equation is obtained whenever all the free algebra generators $x_j \in S$ on the left are replaced by elements of B. The algebras satisfying a fixed set of laws form what is called a *variety* of algebras. A typical law of degree 3, for example, is

$$x_1(x_2x_3) = (x_1x_2)x_3;$$

it determines the variety of associative algebras. (Here, as is usual, we have omitted the outermost brackets.)

2. A Sketch of the Prehistory of Universal Algebra. There are some branches of mathematics to which one may arguably assign a clearly defined starting point. In the case of Universal Algebra what suggests itself as the beginning, at least in a narrow sense, is Garrett Birkhoff's paper [**1**] *On the Structure of Abstract Algebras* which appeared in the Proceedings of the Cambridge Philosophical Society in 1935. Here the definition of a Universal Algebra appears for the first time and the broad outlines of the subject may already be discerned. In particular "abstract algebras are divided by a very simple scheme into self-contained 'species.' Within each species a perfect duality is found between families of formal laws and the families of algebras satisfying them." Here the term "family of algebras of a given species" is used in a technical sense meaning a class closed under taking subalgebras, homomorphic images and direct products—or what is nowadays called a *variety*.

This said, one must point out that undoubtedly the most influential figure in the movement towards abstraction and generality in algebra which culminated in Universal Algebra was Emmy Noether, who died in 1935. In the twenties her school had investigated the notion of a group with operators, which is already very general, including, for example, groups, rings and linear spaces. As Birkhoff himself said in 1946 [**2**] "they had developed many of the most important ideas of Universal Algebra"—for example, the three fundamental isomorphism theorems, which are now recognized as theorems of Universal Algebra, were given for groups with operators in van der Waerden's *Moderne Algebra* [**13**] in 1931—and van der Waerden was of course a protegé of Noether. As Birkhoff in the 1935 paper refers to van der Waerden and also, to Hasse's *Höhere Algebra* [**9**], there is no doubt that he was familiar with the work of the Noether school. Birkhoff's first papers were on lattices and it is not surprising that he should have sought a notion more general than that of a group with operators to express these structures as well.

While it is certainly the sequence Dedekind-Noether-Birkhoff which is the main line of influence in the prehistory of Universal Algebra, I want to consider another very different family of ideas. In 1898 Cambridge University Press published a large and impressive-looking tome [**14**] bearing the title *Universal Algebra, Volume I* by A. N. Whitehead (1861–1947). It was Whitehead's first book, and five years after it appeared he was elected to the Royal Society. Birkhoff has said (in [**3**]) that it was from this book that he borrowed the name "Universal Algebra." Whitehead was Professor of Philosophy at Harvard from 1924 to 1937, a period which includes the years in which Birkhoff was a student there (and his father Professor of Mathematics).

To give some idea of what is in Whitehead's book, let me quote from the preface:

> After the general principles of the whole subject have been discussed in Book I of this volume, the remaining books of the volume are devoted to the separate study of the Algebra of Symbolic Logic and of Grassmann's Calculus of Extension.

In fact, there are seven books in all, of which only one considers symbolic logic; it turns out that even Book I, called *The Principles of Algebraic Symbolism*, is based on Grassmann.

Volume II of Whitehead's book never appeared, perhaps because it was around 1898 that he began his collaboration with a greater mind, Bertrand Russell. It was intended that it would deal with the theory of associative linear algebras which had begun with Benjamin Peirce's famous *Memoir* [**12**] of 1870 (which only appeared in print in 1881).

Perhaps the main thing that one might expect of such a work is unification, and that is what Whitehead aimed for and what one may fairly say, I think, he failed to achieve.

> Universal Algebra [he says] is the name applied to that calculus which symbolizes general operations, defined later, which are called Addition and Multiplication. There are certain general definitions which hold for any process of addition and others which hold for any process of multiplication. These are the general principles of any branch of Universal Algebra.

Briefly these principles amount to commutativity and associativity of addition and distributivity (left and right) of multiplication over addition. These ideas are taken without improvement straight from the first chapter of Grassmann's *Ausdehnungslehre* [**6**] of 1844. Even the idea of considering in one book the algebras of Boole and of Grassmann is not original. Peano had done the same thing in 1888 (in a book [**11**] of which Whitehead knew), and had written in his preface as follows:

> The geometric calculus is preceded by an introduction which treats the operations of deductive logic; these present great analogies with those of algebra and of the geometric calculus.

After distinguishing the algebras of Boole and of Grassmann as being respectively of the nonnumerical genus (meaning that $a + a = a$) and of the numerical genus (meaning that $a + a = 2a \neq a$), Whitehead goes on to consider the two topics in isolation from one another. One need say no more of this first *Universal Algebra*—not really the first even, since Sylvester had previously used the title for a paper on matrices—except to remark that Whitehead (along with all Grassmann's expositors that I know of) offers a presentation of Grassmann's ideas which is inferior to the original.

3. Grassmann and Boole. Let us step back now to the middle of the nineteenth century and the work of Boole (1815–1864) and Grassmann (1809–1877). I do not want to spend much time on the origins of Boole's algebra of logic which first appeared in 1847. It is perhaps worth noting, though, some similarities in the lives and personalities of these two highly original individuals. Both were fascinated by languages, philosophy, and theology, and came to mathematics late (around the age of twenty) and both had the good fortune never to take a university course in mathematics. Both were interested in the teaching of elementary mathematics, and, perhaps most significantly, both fell under the influence of two great mathematical works, the *Mécanique Analytique* (1788) of Lagrange and the *Mécanique Céleste* (1799–1825) of Laplace. As is well known, there was a very considerable difference in the reception which their ideas received.

Boole became part of an English school of mathematicians who were investigating in piecemeal fashion the so-called "laws of symbolic algebra" and one may say, I think, that had he not discovered his algebra of logic someone else would soon have done so. Grassmann's case is different—as is evidenced by the fact that so many of his ideas were rediscovered many decades later on (and by the blank reception they received initially).

Though both Boole and Grassmann were motivated by what we would call modelling problems, they both came early to the view that mathematics deals with formal structures and that its truth does not reside in any interpretation of its symbols; in this they were pioneers. Thus in 1847 in *The Mathematical Analysis of Logic* [**4**] Boole writes that

> ... the validity of the processes of analysis does not depend upon the interpretation of the symbols which are employed, but solely upon the laws of their combination. Every system of interpretation which does not affect the truth of the relations supposed is equally admissible.

And three years earlier in his introduction to the first *Ausdehnungslehre* we find Grassmann expressing the opinion that

> ... geometry can in no way be viewed, like arithmetic or the theory of combinations, as a branch of mathematics; instead, geometry relates to something already given in nature, namely, space. I also had realised that there must be a branch of mathematics which yields in a purely abstract way laws similar to those of geometry, which is limited to space. By means of the new analysis it is possible to form such a purely abstract branch of mathematics; indeed this new analysis, developed without assuming any principles established outside its own domain and proceeding purely by abstraction, was itself this science.

A final link between them is the fact that both were in a sense anticipated by Leibniz (1646–1716) who had sought both a "calculus ratiocinator" (or algebra of logic) and a "characteristica geometrica" (or algebra of geometry), and indeed had developed attempts at both, albeit somewhat more successfully with the logical algebra than with the geometric. In Leibniz's vision there was to be a "characteristica universalis" which would embrace both these algebras; this has indeed come to pass, and one may say perhaps that in a very wide sense it is Leibniz who is the father (or, more accurately, the prophet) of Universal Algebra. Most of these ideas of Leibniz were unpublished until around his bicentenary in 1846, and the question of their influence on Grassmann and Boole is a delicate one which I will not go into.

The fame of Hermann Grassmann today rests on his creation of exterior algebra. What should be realized, though, is that he is in fact the main creator of linear algebra in the modern sense, and that while it is true that geometric considerations motivated his work, he wished to be seen as being an abstract algebraist and that is what he was. In the *Ausdehnungslehre* of 1844 Grassmann plainly wanted to develop his theory in axiomatic "modern algebra" style, but this he was unable to do. To find a modern parallel to the constructive approach which he adopted in the second *Ausdehnungslehre* one must go, I think, to a text in Universal Algebra.

4. Grassmann's Products. I come now to the main part of this paper. It concerns Chapter 2 of Grassmann's *Ausdehnungslehre* of 1862 (which, though it bears the date 1862 on the title page, actually appeared in 1861).

As background we need to know that in the first chapter, which is called *Addition and subtraction of extensive quantities, and their multiplication and division by numbers*, Grassmann had developed in detail, essentially as it is done today, the theory of basis and dimension for finite-dimensional linear spaces. The arena in which this is done is a finite-dimensional real linear space of *extensive quantities*; this is called a *region* and a basis for it, $e_1, e_2, \ldots, e_n$, is called a *system of units*. Although attention is focused on a well-defined space in this way, one must realize that not all extensive quantities are spanned by $e_1, e_2, \ldots, e_n$. One would like to say, in modern fashion, that the extensive quantities form a linear space of which the region spanned by $e_1, e_2, \ldots, e_n$ is a subspace, but Grassmann does not pin down an ambient space in this way.

Grassmann begins Chapter 2 by defining the product of quantities $a = \Sigma\alpha_i e_i$ and $b = \Sigma\beta_j e_j$ from the region determined by $e_1, e_2, \ldots, e_n$ to be

$$ab = \Sigma\alpha_i\beta_j[e_i e_j].$$

After remarking in passing that, being an extensive quantity, this product must itself be a linear combination of a system of units, and that particular kinds of "product structure" will be singled out when one specifies what this system of units is to be and how the products $[e_i e_j]$ are generated by them, he indicates that he will, for the present, "deal only with laws which follow from the general definition of product (above), and which therefore hold for every kind of product." He immediately proves, among other things, that for extensive quantities $a = \Sigma\alpha_i e_i$, $b = \Sigma\beta_j e_j$, $c = \Sigma\gamma_j e_j$ and a real number α one has

$$(a+b)c = ac + bc,\ c(a+b) = ca + cb,$$
$$\alpha(ab) = (\alpha a)b \text{ and } b(\alpha a) = \alpha(ba).$$

In the second part of Chapter 2, called *Products of several quantities*, higher order products like $(ab)((cd)e)$ are considered. They are obtained by iteration of multiplication (juxtaposition) of pairs of extensive quantities; it is clear that there is no assumption of associativity. Since any three quantities a, b and c necessarily belong to some region, the above laws must hold for arbitrary quantities. At this point one may reasonably say that Grassmann's quantities form a free (nonassociative) linear algebra; but again it must be emphasized that this ambient space is not seen as a whole but rather is explored by means of local investigations confined to finitely-generated subalgebras.

The third part of Chapter 2 begins with a definition:

> If a product structure is determined by the fact that some of the products of units are dependent, then I call each equation expressing such a dependence a determining equation for that type of product structure. A set of p determining equations, none of which is derived from the others, and such that there is no other equation expressing dependence among the products, is called a system of determining equations associated with that product structure.

I find it helpful here to think in terms of a multiplication table. In a linear algebra generated by $e_1, e_2, \ldots, e_n$ the multiplicative structure is determined by listing all products of pairs of the linear space generators

$$e_1, e_2, \ldots, e_n, (e_1e_1), (e_1e_2), \ldots, (e_ne_n), e_1(e_1e_1), e_1(e_1e_2), \ldots .$$

In the free case these products are obtained simply by juxtaposition. The effect of a determining equation

$$\Sigma\alpha_j E_j = 0$$

(where the E_j are elements of this sequence of linear space generators) is to allow us to eliminate one element, say E_1, from this list, and indeed also to eliminate any other element which has E_1 as a factor; for example, if $e_2e_1 = -e_1e_2$ is a determining equation, then after deleting the element $(e_3(e_2e_1))(e_4e_5)$ from our list (since it equals $-(e_3(e_1e_2))(e_4e_5)$) we still have a list which spans our algebra. It appears that Grassmann has here delineated, albeit inelegantly, the idea of presenting an algebra by means of generators $(e_1, e_2, \ldots, e_n)$ and relations $(\Sigma\alpha_j E_j = 0)$. (Nowadays we regard this algebra as the quotient of the free algebra by the ideal generated by the $\Sigma\alpha_j E_j$.) Any finitely-generated linear algebra may be obtained in this way.

We now come to a key definition:

> A product structure whose determining equations remain valid when the units occurring in them are replaced by arbitrary quantities spanned by the units is called a linear product structure.

For example (confining attention as usual to products of elements of the region spanned by $e_1, e_2, \ldots, e_n$) what it means for the determining equation

$$e_2e_1 = e_1e_2$$

to be linear is that

$$ba = ab$$

for every quantity $a = \Sigma\alpha_i e_i$ and $b = \Sigma\beta_j e_j$. We are here close to the general concept of a law; indeed the commutative law $x_2x_1 = x_1x_2$ will be satisfied by arbitrary quantities if and only if this linear determining equation holds on every region.

Let me now quote from P. M. Cohn's *Universal Algebra* [**5**]:

> Any systematic study of linear K-algebras would proceed by considering the possible sets of laws.

This is precisely what Grassmann does. He immediately proves the following theorem:

> *For products of two factors there are, apart from the product structure with no determining equations, and the one in which all products are zero, only two types of linear product, namely the one whose system of determining equations has the form*
>
> $$e_ie_j + e_je_i = 0$$
>
> *and the one for which it has the form*
>
> $$e_ie_j = e_je_i,$$
>
> *where both i and j may take values from* 1 *to n, and* $e_1, e_2, \ldots, e_n$ *are units.*

What this means is that if an equation

$$\Sigma\alpha_{ij}x_ix_j = 0$$

is satisfied when the x_i are replaced by arbitrary elements $a = \Sigma\alpha_i e_i$ of the region spanned by $e_1, e_2, \ldots, e_n$, then it must be the case that for all such $a = \Sigma\alpha_i e_i$ and $b = \Sigma\beta_j e_j$, either $ab = 0$ or $ba = ab$ or $ba = -ab$. I would like to consider the relationship of this result to the following theorem which was apparently first stated explicitly (in **[10]**) in 1950:

> *Any law for linear algebras that is homogeneous and of degree 2 and does not hold trivially is equivalent to either $xy = 0$ or $yx = xy$ or $yx = -xy$.*

I claim that this is an immediate consequence of Grassmann's result. Indeed, suppose that we have an algebra in which such a law holds. Grassmann's theorem then entails that for each region the multiplication of pairs of elements is governed by one of the three formulae $xy = 0$, $yx = xy$ or $yx = -xy$. To prove the stated theorem it will suffice to show that for any two regions the governing formula must be the same. The only way that it could happen that none of the three holds identically would be if $yx = xy$ held nontrivially on some region and $yx = -xy$ on some other; and that cannot occur since a single law must hold on the join of the two regions.

One may fairly say, I think, that had Grassmann been able to break the spell which restricted his vision to local aspects of the algebra of quantities, he would have proved this theorem which is evidently fundamental in the study of linear algebras.

It may be argued that the proper job of the historian is to see ideas in the context of their time, but, in their time, Grassmann's ideas were not comprehended. To do justice to this great mathematician (as a creator rather than as an influence) one must, I believe, see his work in the light of twentieth century developments in algebra. In a real sense, he is our contemporary.

This is the manuscript of an invited address given at the First Australian Conference in the History of Mathematics at Monash University in November 1980. It has appeared in the Proceedings of the Conference, edited by John Crossley and published by the Mathematics Dept., Monash University, Clayton, VIC 3168, Australia.

I would like to thank Barry Gardner for many enlightening conversations about Universal Algebra, and Mike Newman for drawing my attention to reference [3].

References

1. Garrett Birkhoff, On the structure of abstract algebras, Proc. Cambridge Philos. Soc., 31 (1935) 433–454.
2. ______, Universal algebra, Canad. Math. Congress Proc., 1946.
3. ______, The rise of modern algebra to 1936, Graduate Studies, no. 13, Texas Tech. University, 1976.
4. George Boole, The Mathematical Analysis of Logic, Cambridge, 1847.
5. P. M. Cohn, Universal Algebra, Harper and Row, New York, 1965.
6. Hermann Grassmann, Die Lineale Ausdehnungslehre, Leipzig, 1844.
7. ______, Die Ausdehnungslehre, Leipzig, 1862.
8. George Grätzer, Universal Algebra, Van Nostrand, Princeton, N.J. 1968.
9. H. Hasse, Höhere Algebra, Berlin, 1927.
10. A. I. Malcev, On algebras with identical defining relations, Mat. Sb. 26 (68) (1950) 19–33.
11. G. Peano, Calcolo Geometrico secondo l'Ausdehnungslehre di H. Grassmann, Torino, 1888.
12. B. Peirce, Linear Associative Algebra, Amer. J. Math., 4 (1881) 97–229.
13. B. L. van der Waerden, Moderne Algebra, 2 vols., Berlin, 1931.
14. A. N. Whitehead, Universal Algebra, vol. 1, Cambridge, 1898.

ON GENERAL PROPERTIES OF QUADRATIC SYSTEMS

CARMEN CHICONE
Department of Mathematics, University of Missouri, Columbia, MO 65211

TIAN JINGHUANG
The Institute of Mathematical Sciences of the Chengdu Branch of the Chinese Academy of Sciences and the Department of Mathematics, Sichuan University, China.

1. Introduction. An nth degree polynomial system of differential equations on the plane is a pair of autonomous differential equations

$$\dot{x} = P_n(x, y) = \sum_{i+k=0}^{n} a_{ik}x^i y^k$$
$$\dot{y} = Q_n(x, y) = \sum_{i+k=0}^{n} b_{ik}x^i y^k \qquad (1_n)$$

where the polynomials P_n and Q_n are relatively prime and at least one of them has degree n. In the special case $n = 2$, which is most important for this exposition, the system is called quadratic. Of course, the basic problem is to find the qualitative structure of the integral curves in the plane of the general system of form (1_n).

The study of polynomial systems has a long history in mathematics and in the physical sciences. Of the many possible physical examples we mention the van der Pol oscillator

$$\dot{x} = y$$
$$\dot{y} = -x - r(x^2 - 1)y$$

which has been a fundamental example in the general development of the qualitative theory of ordinary differential equations [**1**], and the quadratic system

$$\dot{x} = x(a_0 + a_1x + a_2y)$$
$$\dot{y} = y(b_0 + b_1x + b_2y) \qquad (2)$$

which contains several physical models as special cases. In particular, this quadratic system contains the Volterra-Lotka equations [**16**]

$$\dot{x} = x(A - By)$$
$$\dot{y} = y(Cx - D)$$

which are used in mathematical ecology to model the populations of a predator-prey system and are used in chemistry to model the concentrations of two ideal chemical reactants. This ubiquitous set of equations also arises in astrophysics and fluid mechanics. The reader may wish to show [cf. **8**] that the Emden-Fowler equation

$$(\sigma^2\eta')' + \sigma^\lambda\eta^n = 0$$

is transformed by the change of variables

$$x = \frac{\sigma\eta'}{\eta}, \quad \eta = \frac{\sigma^{\lambda-1}\eta^n}{\eta'}, \quad t = \ln|\sigma|$$

to a system of type (2) as is the Blasius equation

$$\eta''' + \eta\eta'' = 0$$

Carmen Chicone is an Assistant Professor at the University of Missouri–Columbia.
Tian Jinghuang is an instructor at Sichuan University.

by the change of variables

$$x = \frac{\eta\eta'}{\eta''}, \quad \eta = \frac{\eta'^2}{\eta\eta''}, \quad t = \ln|y'|.$$

The history of polynomial differential equations is rooted in the natural sciences. However, the subject as pure mathematics has also enjoyed high status since 1900 by virtue of its inclusion in Hilbert's famous list of problems as part of problem 16 [**15**]:

> ... I wish to bring forward a question which, it seems to me, may be attacked by the ... method of continuous variations of coefficients, and whose answer is of ... value for the topology of families of curves defined by differential equations. This is the question as to the maximum number and position of Poincaré's boundary cycles (cycles limites) for a differential equation of the first order and degree of the form
>
> $$\frac{dy}{dx} = \frac{Y}{X},$$
>
> here X and Y are rational integral functions of the nth degree in x and y.

Here a limit cycle is a periodic orbit which is the α or ω limit set of some point not on the periodic orbit. The problem is to bound the number of possible limit cycles of a general system of form (1_n) in terms of n. This problem has proved to be one of the most difficult of Hilbert's entire collection; indeed, it remains unsolved even for quadratic systems [**2**].

We shall have more to say about the 16th problem in the course of our exposition. But, before we begin, we wish to mention two indispensable sources: the survey of Coppel [**8**] which contains an exhaustive bibliography up to 1966 and the beautiful paper of Tung Chin-chu [**24**] which contains many of the results and ideas which we shall explore.

2. Limit Cycles. As we mentioned in the introduction, the investigation of the limit cycles for polynomial systems remains the most important and difficult question in the subject. Before presenting the theorems of Tung Chin-chu we review some aspects of the history of Hilbert's problem.

Recall that we seek to bound the number of limit cycles for (1_n) in terms of the degree n. This immediately brings up the question: Does a bound exist? In fact, it is unknown whether or not a finite bound exists. However, we do have the theorem of Dulac [**10**] which states that, given a *particular choice* of coefficients for a system of form (1_n), then the particular differential equation has at most a finite number of limit cycles. Dulac's paper is very long (over 100 pages) and notoriously difficult to read. Hence, we propose the following problem.

PROBLEM 2.1. Give a modern exposition of Dulac's theorem and, if possible, find a shorter proof.

Dulac's solution is based on the following idea. Assume first that there are infinitely many limit cycles and that they accumulate on a periodic orbit O. There is a local section Σ at any point p of O and Dulac considers the Poincaré return map $P\colon \Sigma \to \Sigma$. Since the polynomial vector field is analytic, it follows that P is analytic and that the function $f(x) = P(x) - x$, which is also analytic, has a sequence of zeros with a limit point. This is a contradiction. However, the infinite collection of limit cycles may accumulate on a separatrix cycle which consists of saddle points and connecting separatrices. In this case one can still find a local Poincaré section and section map P, but now P may not even be differentiable so the same argument can't be made. Dulac argues that nonetheless $f(x) = P(x) - x$ can't have an accumulation point of its zeros. However, this is the part of the argument which is so difficult to understand. Possibly then the validity of Dulac's argument should remain in doubt until a new argument is found, or until someone solves 2.1.

A less well-known result due to Diliberto [**9**] actually does give a bound for the number of limit cycles at the expense of making a stronger assumption. To understand Diliberto's result, recall that the divergence of a vector field given by

$$\dot{x} = P(x, y)$$
$$\dot{y} = Q(x, y)$$

is $\operatorname{div}(P, Q) = \partial P/\partial x + \partial Q/\partial y$.

A periodic orbit γ is called hyperbolic (exponentially attracting or repelling) if

$$\int_\gamma \operatorname{div}(P, Q)\, dt$$

is not zero. If $\operatorname{div}(P, Q) > 0$ (respectively $\operatorname{div}(P, Q) < 0$) at all points of γ then γ is called strongly stable (respectively strongly unstable).

THEOREM 2.2. *If all periodic solutions of* (1_n) *are either strongly stable or strongly unstable, the total number of periodic solutions is less than*

$$\tfrac{1}{2}(n-2)(n-3)+1.$$

If all periodic orbits are nested and surround a single critical point, the estimate can be sharpened to the greatest integer less than or equal to $(n-1)/2$.

Idea of the proof. One shows that there are at least as many closed finite branches of $\operatorname{div}(P, Q) = 0$ as there are periodic orbits by observing that in a nest of periodic orbits adjacent periodic orbits alternate between being strongly stable and strongly unstable. Then one uses the classical result of Harnack [cf. **2**] that an algebraic curve of degree n has at most

$$1 + \frac{(n-1)(n-2)}{2}$$

ovals in the plane and $[n/2]$ ovals when they are nested.

It is interesting to note that Harnack's Theorem is the subject of the first part of Hilbert's problem, i.e., to describe the ovals of an algebraic plane curve. Possibly Diliberto's Theorem is a realization of Hilbert's intuition and the reason for his linkage of the two parts of the 16th problem.

Another result which is directly related to the 16th problem was proved by Bautin [**3**]. We say that a finite set of limit cycles $\{\gamma_k\}$, $k = 1, \ldots, m$ disappear into an equilibrium point for a parametrized system

$$\dot{x} = P(x, y, \sigma)$$
$$\dot{y} = Q(x, y, \sigma)$$

if there is a path $\sigma(\tau)$ for $0 \leqslant \tau \leqslant 1$ with all limit cycles surrounding a single equilibrium such that as $\tau \to 1$ no two cycles merge and diameter $L_k(\sigma(\tau)) \to 0$. Bautin's Theorem states that for a quadratic system at most three limit cycles can disappear into an equilibrium. Again, Bautin's theorem is rather difficult to prove. For this bifurcation problem on the plane we give

PROBLEM 2.3. Find a short proof of Bautin's theorem.

Perhaps inspired by Bautin's theorem, Petrovskiĭ and Landis [**19**], [**20**] in a series of papers proposed a solution to Hilbert's problem. In fact, they asserted that in the quadratic case the maximum number of limit cycles is three. Their work, which incorporates ideas from the algebraic geometry of complex projective space, has been questioned since its publication. They published an errata [**20**], but were unable to correct all of the mistakes that were found. Unfortunately, due to these errors the ideas contained in this work have not been widely used by other mathematicians. Since most researchers in differential equations are not trained in algebraic geometry, the eventual fate of these ideas may well remain unknown until they are revived by geometers.

All hope that, in fact, three is the correct bound for the quadratic case was removed in 1980 by

the work of the Chinese mathematician, Shi Songling [22]. He proved that the system

$$\dot{x} = \lambda x - y - 10x^2 + (5 + \delta)xy + y^2$$
$$\dot{y} = x + x^2 + (8\varepsilon - 25 - 9\delta)xy$$

where

$$\delta = -10^{-13}, \quad \varepsilon = -10^{-52} \quad \text{and} \quad \lambda = -10^{-200}$$

has at least four limit cycles!

Recall that Hilbert asked not only for the maximum number but also for the position of the limit cycles. In an attempt to answer this question, we come to the work of Tung Chin-chu. Although he accepted the results of Petrovskiĭ and Landis (he did not rely on their methods), his results are valid independent of the number of limit cycles. In the remainder of this section we shall present, with some modifications, the basic results of his theory.

We denote the general quadratic system by

$$\begin{aligned} \dot{x} &= P(x, y) \\ \dot{y} &= Q(x, y). \end{aligned} \tag{2.4}$$

Also, we say that (2.4) has a contact at the point (x_0, y_0) of the line L if the vector $(P(x_0, y_0), Q(x_0, y_0))$ is parallel to L.

THEOREM 2.5. *The system* (2.4) *has the following properties*:

(a) *there are at most four critical points*,

(b) *three critical points are never collinear*,

(c) *if* L: $ax + by + c = 0$ *is not an invariant set*, *then there are at most two critical points or contacts on* L. *Furthermore*, *if* L *contains two critical points or points of contact*, *the orientation of the vector field* $X = (P, Q)$ *on the infinite segments cut off by the critical points or points of contact is opposite to the orientation of* X *on the finite segment*, *and*

(d) *a line joining two critical points is an isocline*.

Proof. (a) follows from the fact that two conics meet in at most four points. If L is a line containing three critical points, then L meets each conic $P(x, y) = 0$ and $Q(x, y) = 0$ in three points. This is possible only if L is a factor of both P and Q contrary to the fact that P and Q are relatively prime. This proves (b).

Assume L is not invariant. The critical points and points of contact of X along L are solutions of the equations

$$\begin{aligned} aP + bQ &= 0 \\ ax + by + c &= 0. \end{aligned}$$

As the solution set is the intersection of a line and a conic with no common factor, there are at most two solutions. If there are two solutions, the finite segment cut off on L lies in the region $aP + bQ < 0$ and the two infinite segments lie in the region $aP + bQ > 0$. This proves (c).

For (d) assume L contains two critical points of X and without loss of generality that one of these points lies at the origin. L is given by $y = mx$ and the slope of X along L is

$$\text{slope} = \frac{Q(x, mx)}{P(x, mx)}.$$

As $Q(x, mx)$ and $P(x, mx)$ are quadratic polynomials in one variable with equal roots, their quotient is a constant. This proves (d).

For the nth degree case one can show (a) there are at most n^2 critical points (Bézout's Theorem), (b) $n + 1$ critical points are never collinear, and (c) there are at most n critical points and contacts on a noninvariant line.

THEOREM 2.6. *A periodic orbit of a quadratic system encloses a strictly convex region.*

Proof. If the region R enclosed by the periodic orbit is not convex, there are two points in R such that the line segment between them contains a point exterior to the periodic orbit. This implies that the line determined by the two points contains at least three contacts contrary to 2.5 part (c). Also, 2.5 part (c) is violated if the periodic orbit contains a line segment.

This theorem is false for cubic systems. For example, the van der Pol equation

$$\dot{x} = y$$
$$\dot{y} = -x - r(x^2 - 1)y$$

which has a unique limit cycle with x-extremities of magnitude greater than $\sqrt{3}$ around the origin may fail to have a limit cycle which encloses a convex region. For this one computes

$$\frac{dy}{dx} = -r(x^2 - 1) - \frac{x}{y}$$

and

$$\frac{d^2y}{dx^2} = -2rx - \frac{1}{y} + \frac{x}{y^2}\frac{dy}{dx}.$$

At the point $(-1, y)$ in the half-plane $y > 0$ we have $dy/dx = 1/y > 0$ and

$$\frac{d^2y}{dx^2} = 2r - \left(\frac{1}{y} + \frac{1}{y^3}\right).$$

Hence, when the positive number r is sufficiently large (such that $d^2y/dx^2 > 0$) the limit cycle encloses a nonconvex region.

We obtain as a corollary

THEOREM [**26**] 2.7. (a) *A periodic orbit of a quadratic system has unique points with maximum abscissa and minimum abscissa, maximum ordinate and minimum ordinate.* (b) *Exactly one branch of the curve $Q(x, y) = 0$ passes through the points with maximum ordinate and minimum ordinate and exactly one branch of $P(x, y) = 0$ passes through the points with maximum and minimum abscissa.*

Proof. Part (a) follows from the strict convexity of the region enclosed by the periodic orbit. Note that at the points with maximum and minimum ordinate we have $dy/dx = 0$. Hence Q vanishes at these points. Of course, P does not vanish at these points because they are not critical points. Also, Q vanishes at only these points by their uniqueness. Now, if the points lie on distinct branches of the curve $Q = 0$, then the curve $Q = 0$ contains no points in the region interior to the periodic orbit. For, if it did, it would have to contain at least one other point on the periodic orbit which is not one of the original points. But this leads to a contradiction since any closed orbit contains at least one critical point in the region it bounds and a branch of $Q = 0$ must contain this critical point.

THEOREM 2.8. *A periodic orbit of a quadratic system surrounds a unique critical point which is elementary and has index* $+1$.

Proof. It follows from 2.7(b) that in the interior of a periodic orbit γ there is a unique intersection of one branch of the curve $Q(x, y) = 0$ and one branch of the curve $P(x, y) = 0$, i.e., a unique critical point. At the intersection the normal vectors

$$\left(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}\right) \quad \text{and} \quad \left(\frac{\partial Q}{\partial x}, \frac{\partial Q}{\partial y}\right)$$

of both curves evidently cannot coincide, and the determinant

$$\begin{vmatrix} \dfrac{\partial P}{\partial x} & \dfrac{\partial P}{\partial y} \\ \dfrac{\partial Q}{\partial x} & \dfrac{\partial Q}{\partial y} \end{vmatrix} \neq 0$$

at the critical point. Hence, it is elementary.

The change in the angle measured anticlockwise from the positive direction of the x-axis to the direction of the vector (P, Q) along the positively oriented closed curve γ after a full cycle is $+2\pi$. Hence, by the definition of the index of a closed curve and a critical point we have ind $S = \operatorname{ind}\gamma = +1$.

Using these theorems, especially 2.5, we can easily determine the possible positions of the limit cycles for a quadratic system.

THEOREM 2.9. *The positions of the limit cycles of a quadratic system are as follows*:
(a) *there are no limit cycles*,
(b) *there is exactly one limit cycle or one nest of limit cycles all having the same orientation*, *or*
(c) *there are exactly two distinct limit cycles or two nests of limit cycles with no points common to the bounded regions cut off by limit cycles in distinct nests and such that one nest has all orbits oriented clockwise and the other nest has all orbits oriented counterclockwise.*

Proof. Suppose that there are three limit cycles γ_i, $i = 1,2,3$, which cut off mutually disjoint bounded regions. Let p_i be the unique critical point surrounded by γ_i and consider the line segments $[p_1, p_2]$, $[p_2, p_3]$ and $[p_1, p_3]$. The critical points are not collinear so the segments form a triangle. Since the segment $[p_1, p_2]$ contains two critical points, there are no contacts or critical points on the open segment (p_1, p_2) and, hence, γ_1 and γ_2 must have the opposite orientation. (This proves that if two distinct nests exist, they are oppositely oriented.) Similarly, γ_2 and γ_3 are oppositely oriented. But, this implies that γ_1 and γ_3 have the same orientation and we reach a contradiction.

The rest of the argument is similar and is left as an interesting exercise for the reader.

The configurations in Fig. 1 were shown to exist in concrete quadratic systems for (a) by Frommer [**13**], for (b) and (c) by Bautin [**3**], for (d) by Yeh Yen-chien [**26**] and Tung Chin-chu [**24**], for (e) by Tung Chin-chu and for (f) by Shi Songling [**22**]. We point out that in cases (e) and (f) the respective authors really show that there are *at least* as many limit cycles as shown. This is evident in the work of Shi Songling and implicit in the work of Tung Chin-chu (he accepted that the maximum number of limit cycles was three).

PROBLEM 2.10. Find in Tung Chin-chu's example [**24**, p. 873] the actual number of limit cycles. The equations are

$$\dot{x} = P(x, y)\cos\theta - Q(x, y)\sin\theta$$
$$\dot{y} = P(x, y)\sin\theta + Q(x, y)\cos\theta$$

where

$$P(x, y) = xy$$
$$Q(x, y) = -\tfrac{1}{3}(x-1)(x+2) + \tfrac{1}{2}y^2 + \tfrac{1}{3}xy - \tfrac{1}{3}y$$

and θ is a small negative number.

PROBLEM 2.11. Make computer plots of all the known examples. This problem is difficult because the coefficients vary so much in magnitude, but it would be very interesting to have pictures of the examples.

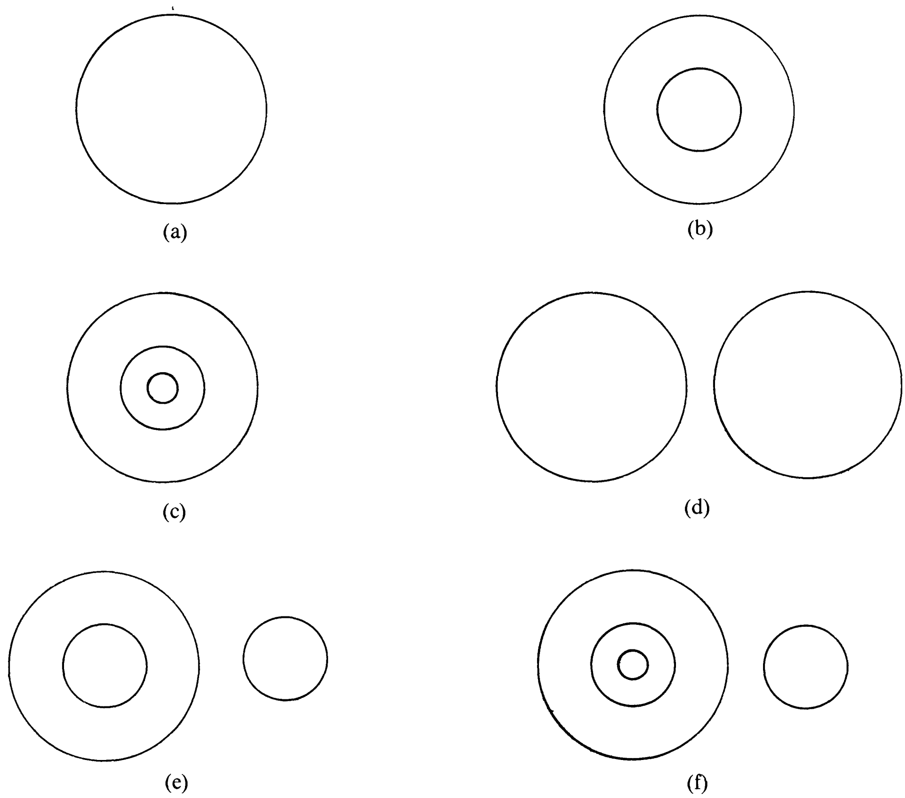


FIG. 1. The "known" configurations of limit cycles in the quadratic case.

3. More About Critical Points. Recall that a center is a critical point p such that some neighborhood of p contains only periodic orbits. A focus is a spiral source or a spiral sink. Clearly, a center or a focus has an orientation.

THEOREM 3.1. *If* (1_n) *has* n *collinear orientable critical points (foci and centers), then two adjacent critical points have opposite orientation. In particular,* **[8]** *if a quadratic system has a focus and a center, they are oppositely oriented. Moreover, the total number of foci and centers for a quadratic system is at most two* **[5]**.

Proof. Argue as in the proof of 2.9 using Tung Chin-chu's theorem 2.5.

The next theorem shows that a periodic orbit of a quadratic system does not surround a node, i.e., a critical point p of (2.4) such that the linear part of (P, Q) at p has two positive (two negative) real eigenvalues.

THEOREM 3.2. *Assume* p *is a critical point of* (2.4) *and that the eigenvalues of the linear part of the vector field at* p *are real. Then, either there is an invariant line through* p *or there is a line* L *through* p *such that* L *contains no other contacts or critical points and the orientation of the vector field along* L *does not change sign at* p.

Proof. By a change of coordinates we may assume $p = (0,0)$. Let L denote the line given by the equation $ax + y = 0$ and consider the orientation of $(\dot{x}, \dot{y})$ along L, i.e., the sign of the expression

$$a\dot{x} + \dot{y} = \sum_{i+j=1}^{2} \left(aa_{ij} + b_{ij}\right)x^i y^j$$

$$= \sum_{i+j=1}^{2} (aa_{ij} + b_{ij})x^i(-ax)^j$$
$$= f(a)x + g(a)x^2$$

where

$$f(a) = -a^2a_{01} + a(a_{10} - b_{01}) + b_{10}$$

and

$$g(a) = b_{20} + a(a_{20} - b_{11}) + a^2(b_{02} - a_{11}) + a^3a_{02}.$$

Since the eigenvalues of the linear part at $(0,0)$ are real, the discriminant of the characteristic equation is greater than or equal to zero, i.e.,

$$(b_{01} - a_{10})^2 + 4a_{01}b_{10} \geqslant 0.$$

This implies that the discriminant of f, as a quadratic in a, is positive and, hence, that f has a real root a_1. There are now two cases. If $g(a_1) = 0$, then

$$a_1\dot{x} + \dot{y} \equiv 0$$

and L is an invariant line through $(0,0)$. If $g(a_1) \neq 0$ we have

$$a_1\dot{x} + \dot{y} = g(a_1)x^2$$

and L satisfies the second conclusion of the theorem.

We can now prove the promised corollary.

THEOREM 3.3. *A periodic orbit of a quadratic system surrounds either a center or a focus. A limit cycle surrounds a focus.*

Proof. It follows from 2.8 and 3.2 that the first conclusion is valid.

We know from [**8**, p. 295] that a limit cycle and a center cannot coexist. Therefore, a limit cycle cannot surround a center. This proves 3.3.

(*Remark.* The second conclusion is one of the most important results about quadratic systems. Theorem 2.9 is easily proved from this conclusion and 3.1.)

We leave as an exercise using 2.5 the

THEOREM 3.4. *If p is a critical point of a quadratic system surrounded by a periodic orbit, then the critical point is oriented and has the same orientation as the periodic orbit.*

PROBLEM 3.5. A cubic system may have a node surrounded by a periodic orbit. In fact this is true for the van der Pol oscillator with an appropriate choice of the parameter.

As we have seen, a quadratic system has at most four critical points. When four critical points occur, they satisfy the beautiful theorem of Berlinskiĭ [**5**].

THEOREM 3.6. *If a quadratic system has exactly four critical points, they form a quadrilateral. If this quadrilateral is convex, then two opposite critical points are saddles and the other pair of critical points are not saddles (i.e., nodes, foci or centers). If the quadrilateral is not convex, then the three exterior vertices are saddles and the interior vertex is not a saddle or, vice versa, the exterior vertices are all not saddles and the interior vertex is a saddle. In particular, a quadratic system can have at most three saddles and at most three nonsaddles.*

Proof. A simple elementary proof due to Kukles and Casanova [**17**] is presented by Coppel [**8**, p. 299]. The reader may wish to repeat this proof. Start by making a linear change of variables so that the critical points are located at $(0,0)$, $(0,1)$, $(1,0)$ and (α, β) where $\alpha \neq 0$, $\beta \neq 0$ and

$\alpha + \beta \neq 1$. The differential equations assume the form

$$\dot{x} = a_1 x(x-1) + b_1 y(y-1) + c_1 xy$$
$$\dot{y} = ax(x-1) - by(y-1) + cxy.$$

Then, examine the Jacobian at the critical points.

4. Additional Properties and Some Recent Results. A classical problem in phase plane analysis is the determination of a center. In particular, if the linear part of the vector field at a critical point p has purely imaginary eigenvalues, is the critical point p in a neighborhood where every orbit is periodic? Although this problem is very difficult, it has been solved completely in the quadratic case.

THEOREM 4.1. *The equation* (*see* [**8**] *for detailed reference*)

$$\frac{dy}{dx} = -\frac{x + ax^2 + (2b+\alpha)xy + cy^2}{y + bx^2 + (2c+\beta)xy + dy^2}$$

has a center at the origin if and only if one of the following holds:

I. $a + c = b + d = 0$.
II. $\alpha(a+c) = \beta(b+d)$, $a\alpha^3 - (3b+\alpha)\alpha^2\beta + (3c+\beta)\alpha\beta^2 - d\beta^3 = 0$.
III. $\alpha + 5(b+d) = \beta + 5(a+c) = ac + bd + 2(a^2+d^2) = 0$.

Another complete success has been achieved for the problem of classification when the system is homogeneous, i.e.,

$$\dot{x} = ax^2 + bxy + cy^2$$
$$\dot{y} = a_1x^2 + b_1xy + c_1y^2.$$

We refer the reader to [**18**], [**25**] for the proofs and to [**8**] for additional historical references.

A related problem is to classify all quadratic systems with no limit cycles (avoiding Hilbert's problem). These results are due to Gonzalez [**14**], Tavares [**23**], and dos Santos [**11**]. The paper of dos Santos contains historical comments and a complete set of references. In this context another special property of quadratic systems was found in [**6**]. In particular, suppose the quadratic system is obtained as the gradient of a cubic polynomial potential and that the gradient has two saddle points connected by an orbit. Then this connecting orbit is part of an invariant straight line.

PROBLEM 4.2. Find a short proof of this theorem.

A number of authors have investigated quadratic systems with special integrals. Possibly the most beautiful result in this direction is the example of Chin Yuan-shun [**7**] who showed that a quadratic system can have an ellipse as a limit cycle. In addition to the references on this topic in [**8**], we also cite [**12**] as a typical recent example.

In such investigations it is often convenient to assume that the general quadratic system can be expressed in a canonical form, i.e., as a quadratic system topologically equivalent to the original system but with fewer nonzero coefficients. A very useful canonical form is given by

$$\dot{x} = -y + \delta x + lx^2 + mxy + ny^2$$
$$\dot{y} = x + ax^2 + bxy.$$

This canonical form can be obtained if and only if the original quadratic system has an elementary critical point with index $+1$. To see this, note first that the canonical form has the origin as an elementary critical point with index $+1$. On the other hand, under our assumption one can obtain the canonical form

$$\dot{x} = k + dx + ey + lx^2 + mxy + ny^2$$
$$\dot{y} = ax^2 + bxy + \alpha x$$

as in [27]. Translation of the critical point to the origin yields the form (with reassigned values of the coefficients)

$$\dot{x} = dx + ey + lx^2 + mxy + ny^2$$
$$\dot{y} = \alpha x + ax^2 + bxy.$$

Since the critical point at the origin is assumed to have index $+1$, the determinant of the linear part is a positive number D^{-1}. Now, multiply the vector field by D and note that this is simply a reparametrization of the integral curves and, hence, does not change the topological type. After this, the rescaled equations have linear part with determinant equal to 1. Hence, a linear change of coordinates will result in the desired canonical form. From 2.8 or 3.3 we know that all quadratic systems having at least one limit cycle must be contained in this canonical form. The system having at least four limit cycles which was given by Shi Songling [22] is found from this canonical form.

THEOREM 4.3. *For a quadratic system having at least three critical points, a center and a focus cannot coexist.*

Proof. 4.3 follows from Theorem 2, [5] and Theorem 3, [28] of Berlinskiĭ.

A very useful theorem for phase plane analysis in general is the Bendixson–Dulac criterion for the absence of limit cycles. Let B be a positive function defined on a simply connected plane region R and let

$$\begin{aligned} \dot{x} &= f(x, y) \\ \dot{y} &= g(x, y) \end{aligned} \tag{4.3}$$

be a differential system defined on R.

THEOREM 4.4. (a) *If* $\operatorname{div}(Bf, Bg)$ *does not vanish on* R, *then* 4.3 *has no periodic orbits in* R. (b) *If* $\operatorname{div}(Bf, Bg) \equiv 0$, *then there are no limit cycles in* R. (c) *If* A *is an annular region on which* $\operatorname{div}(Bf, Bg)$ *does not vanish, then* A *contains at most one limit cycle.*

Proof. The system (4.3) has a limit cycle if and only if

$$\dot{x} = Bf(x, y) = P(x, y)$$
$$\dot{y} = Bg(x, y) = Q(x, y)$$

has a limit cycle γ. Let D be the bounded region bounded by γ and compute

$$\begin{aligned} 0 &= \int_\gamma \dot{x}\dot{y} - \dot{y}\dot{x} \\ &= \int_\gamma P dy - Q dx \\ &= -\int\int_D \operatorname{div}(P, Q)\, dx\, dy \end{aligned}$$

to see (a). To prove (b) observe that, if γ is a limit cycle, the points on one side of γ, say the outside, spiral toward γ (or they all spiral away from γ). There is a simple closed curve C outside γ such that $(\dot{x}, \dot{y})$ points toward γ at all points of C. Consider the annulus A formed by C and γ and let N be the outward normal field. Then

$$\int_{\partial A} (P, Q) \cdot N ds \neq 0$$

but,

$$\int_{\partial A}(P,Q)\cdot N ds=\int\int_A \operatorname{div}(P,Q)\,dx\,dy.$$

The proof of (c) is similar and is left to the reader.

PROBLEM 4.5. Prove Bautin's Theorem [**4**]: The system

$$\dot{x}=x(a_0+a_1x+a_2y)$$
$$\dot{y}=y(b_0+b_1x+b_2y)$$

has no limit cycles.

(Hint: Find an appropriate choice for $B(x,y)$.) This fact seems to be less well known than its importance would indicate. See for example the discussion in [**16**, pp. 264–265]. Also note that the system can have periodic orbits. In fact, show that all orbits of the Volterra–Lotka model are periodic.

PROBLEM 4.6. Consider the system

$$\dot{x}=ax+by+rf(x,y)$$
$$\dot{y}=cx+dy$$

where $f(x,y)$ is either x^2, xy or y^2. Show that such a system can have periodic orbits. Can such a system have a limit cycle? What if $f(x,y)$ is a single nth degree term?

PROBLEM 4.7. Solve Pugh's Problem. (See [**21**] for a nonelementary version of this problem.) Let

$$\dot{x}=x$$
$$\dot{y}=P(x,y)$$

where P is a polynomial. Study solutions in the infinite strip of points with x-coordinate in $[0,1]$. We call a solution P-periodic if its intersections with the lines $x=0$ and $x=1$ have the same height. Does there exist a bound for the number of P-periodic solutions in terms of the degree of P?

PROBLEM 4.8. With reference to 2.2 show that a limit cycle of a quadratic system can never be strongly stable.

PROBLEM 4.9. As in 4.2, saddle connections of quadratic gradient systems are special; they are line segments. Are saddle connections of nth degree polynomial gradient systems special?

PROBLEM 4.10. A quadratic system has a unique critical point inside any periodic orbit and this point must be a focus or a center. What properties hold for the critical points surrounded by a limit cycle of a 3rd degree system? An nth degree system?

PROBLEM 4.11. A Reeb component of a flow in the plane is a region R in the plane foliated by the flow which is topologically equivalent to the flow of

$$\dot{x}=x^2-1,\qquad \dot{y}=x$$

in the strip$\{(x,y)\mid -1\leqslant x\leqslant 1\}$. Can a quadratic system have two distinct Reeb components? How many Reeb components can exist for an nth degree equation?

PROBLEM 4.12. Does there exist a quadratic system which has a limit cycle surrounded by a separatrix cycle? What about a polynomial system of degree n? An analytic system?

References

1. A. Andronov et al., Theory of Bifurcation of Dynamical Systems on the Plane, Halstead Press, New York, 1973.

2. V. Arnold, Real algebraic geometry (the 16th Hilbert problem), in Proceedings of Symposia in Pure Mathematics, no. 28, Amer. Math. Soc., 1976, pp. 50–51.

3. N. Bautin, On the number of limit cycles which appear with the variation of coefficients from an equilibrium position of focus or center type, Amer. Math. Soc. Transl. No. 100, 1954.

4. N. Bautin, On periodic solutions of a system of differential equations (Russian), Prikl. Math. Meh., 18 (1954) 128.

5. A. Berlinskiĭ, On the coexistence of singular points of different types (Russian), Izv. Vysš. Učebn. Zaved. Matematika, no. 5, 18 (1960) 27–32.

6. C. Chicone, Quadratic gradients on the plane are generically Morse-Smale, J. Differential Equations, 33 (1979) 159–166.

7. Chin Yuan-shun, On algebraic limit cycles of degree 2 of the differential equation $dy/dx = \Sigma_{0 \leq i+j \leq 2} a_{ij} x^i y^j / \Sigma_{0 \leq i+j \leq 2} b_{ij} x^i y^j$, Sci. Sinica, 7 (1958) 934–945.

8. W. Coppel, A survey of quadratic systems, J. Differential Equations, 2 (1966) 293–304.

9. S. Diliberto, On systems of ODE's, Contrib. to the theory of nonlinear oscillations, 1 (1950) 1–38.

10. H. Dulac, Sur les cycles limites, Bull. Soc. Math. France, 51 (1923) 45–188.

11. G. dos Santos, Classification of generic quadratic vector fields with no limit cycles, Lecture Notes in Math., no. 597, 1977, pp. 605–640.

12. R. Evdokimenko, Behavior of integral curves of a dynamic system, Differential Equations, 12 (1976) 1095–1103.

13. M. Frommer, Über das Auftreten von Wirbeln und Strudeln in der Urngebung rationaler Unbestimmtheitsstellen, Math. Ann., 109 (1934) 395–424 (411–414).

14. E. Gonzalez Velasco, Generic properties of polynomial vector fields at infinity, Trans. Amer. Math. Soc., 143 (1969) 201–222.

15. D. Hilbert, Mathematical problems, Mary Newton Transl., Bull. Amer. Math. Soc., 8 (1902) 437–479.

16. M. Hirsch and S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, New York, 1974.

17. I. Kukles and M. Casanova, On the distribution of critical points of the first and second group (Russian), Izv. Vysš. Učebn. Zaved. Matematika, 6 (1964) 88–97.

18. T. Newton, Two dimensional homogeneous quadratic differential systems, SIAM Rev., 20 (1978) 120–138.

19. I. Petrovskiĭ and E. Landis, On the number of limit cycles of the equation $dy/dx = P(x, y)/Q(x, y)$, where P and Q are polynomials of the second degree, Amer. Math. Soc. Transl. Ser. 2, 10 (1958) 177–221.

20. I. Petrovskiĭ and E. Landis, Corrections to the articles "On the number of limit cycles of the equation $dy/dx = P(x, y)/Q(x, y)$, where P and Q are polynomials of the second degree" and "On the number of limit cycles of the equation $dy/dx = P(x, y)/Q(x, y)$, where P and Q are polynomials," Mat. Sb., 48 (90) (1959) 253–255.

21. S. Shahshahani, Periodic solutions of polynomial first order differential equations, Nonlinear Anal., 5 (1981) 157–165.

22. Shi Song-ling, A concrete example of the existence of four limit cycles for plane quadratic systems, Sci. Sinica, 23 (1980) 153–158.

23. M. Tavares Camacho, Generic properties of homogeneous vector fields of degree two in R^3, An. Acad. Brasil. Ciênc., 51 (1979) 31–33.

24. Tung Chin-chu, Positions of limit cycles of the system

$$\frac{dx}{dt} = \sum_{0 \leq i+k \leq 2} a_{ik} x^i y^k, \qquad \frac{dy}{dt} = \sum_{0 \leq i+k \leq 2} b_{ik} x^i y^k,$$

Sci. Sinica, 8 (1959) 151–171, or Chinese J. Math., 8 (1966) 854–874.

25. N. Vulpe and K. Sibirskiĭ, Geometric classification of quadratic differential systems, Differential Equations, 13 (1977) 548–556.

26. Yeh Yen-chien, Periodic solutions and limit cycles of certain nonlinear differential systems, Sci. Record N.S., 1 (1957) 391–394.

27. Yeh Yen-chien et al., A qualitative study of the integral curves of the differential equation

$$\frac{dy}{dx} = \frac{q_{00} + q_{10}x + q_{01}y + q_{20}x^2 + q_{11}xy + q_{02}y^2}{p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy + p_{02}y^2},$$

I, Chinese J. Math., 3 (1963) 1–18.

28. A. Berlinskiĭ, On the behavior of the integral curves of a differential equation (Russ.) Izv. Vysš. Učebn. Zaved. Matematika, no. 2 (15) (1960) 3–18.

Four faces of analysis. Their names are on page 214.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by July 31, 1982. The solver's full post-office address should be on each sheet.

6380. *Proposed by Juris Steprāns, University of Toronto.*

Let k be the kth finite ordinal, and let l be either a finite ordinal k or the first infinite ordinal ω. Let ${}^{\omega}l$ denote the set of functions from ω to l. Is the proposition P:

If $F \in {}^{\omega}l$ and card $(F) < 2^{\aleph_0}$, then there exists $g \in {}^{\omega}l$ such that for all $f \in F$ there are infinitely many m for which $f(m) = g(m)$

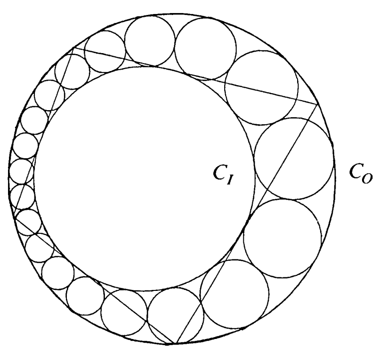
true, false, or independent of ZFC if l is finite?

6381. *Proposed by W. W. Meyer, General Motors Research Laboratories.*

M. D. Fox [1] defines a (Steiner) chain as a sequence of circles each touching its two neighbors and two given boundary circles C_O, C_I. Enlarging on this, with the hypothesis that C_O surrounds C_I, we define a linear chain as a polygon circumscribed by C_O and circumscribing C_I. Linear or circular, a chain is said to have period n/m if it closes on itself, the first link and the $(n+1)$th link coinciding, after m cycles around C_O. Prove that a linear chain of rational period p and a circular chain of rational period q coexist iff

$$p > 2, \quad q > 2, \quad \cos\frac{\pi}{2p}\cos\left(\frac{\pi}{4} - \frac{\pi}{2q}\right) \leq \cos\frac{\pi}{4}$$

and then the boundary circles are uniquely determined as to relative size and eccentricity. (The figure illustrates the case $p = 4$, $q = 19$.)



1. M. D. Fox, Formulae for the curvatures of circles in chains, this MONTHLY, 87 (1980) 708–715.

6382. *Proposed by Hing-kam Lam, The Chinese University of Hong Kong.*

Evaluate $\int_0^1 (\ln x)[\ln(1-x)]^2\, dx/x$.

ANSWERS TO "PHOTOS" ON PAGE 179

Top left: Emile Borel; top right: M. G. Krein; bottom left: S. L. Sobolev; bottom right: E. T. Whittaker.

MEAN CURVATURE, THE LAPLACIAN, AND SOAP BUBBLES

ROBERT C. REILLY
Department of Mathematics, University of California, Irvine, CA 92717

Dedicated to my mother, Anna M. Reilly

1. Introduction. The notion of the mean curvature of a surface is one of the most fruitful in classical differential geometry and is the subject of many research papers. It is treated in all the standard textbooks (see, for example, [**2**], [**6**]), but usually as a special topic embedded in a general discussion of fundamental forms, Codazzi equations, and Gauss maps. (An exception is the classic exposition by Hilbert and Cohn-Vossen [**4**].) In this paper I develop the concept of mean curvature in a simple manner, using only well-known facts from calculus and avoiding the complicated terminology of differential geometry. I point out how this concept is related to the Laplace operator $\Sigma_{j=1}^{3}(\partial^2/\partial x_j^2)$ and I exploit this relation to prove the well-known "Soap-Bubble Theorem" of A. D. Aleksandrov.

This paper grew from a talk I presented to the UCI Math Club; I have tried to make the exposition simple enough that undergraduate math majors could read it with profit. I wish to thank my colleague Mike Fried for his useful comments.

2. The Mean Curvature of a Surface.

(A) *Surfaces.*

DEFINITION. Suppose that S is a connected nonempty subset of $\mathbb{R}^3$. We say that S is a *surface* if on some neighborhood U of S in $\mathbb{R}^3$ there is a C^1 function $h: U \to \mathbb{R}$ such that if $p \in S$ then (i) $h(p) = 0$ and (ii) grad $h(p) \neq 0$. If, for some integer $k \geqslant 1$, h is a C^k function, then we say that S is a C^k *surface.*

REMARKS.

(a) Condition (ii), combined with the Implicit-Function Theorem, ensures that each point of S has a neighborhood (in S) which is the graph of a C^k function of two variables.

(b) There is a tangent plane T_p and a normal line l_p at each point p of S:

$$T_p = \{q \in \mathbb{R}^3 : (q-p) \perp \operatorname{grad} h(p)\} \quad \text{and} \quad l_p = \{q \in \mathbb{R}^3 : (q-p) \| \operatorname{grad} h(p)\}.$$

(c) The orientation of S is determined by singling out one of the two unit vector fields normal to S, grad $h/|$grad $h|$ and $-$ grad $h/|$grad $h|$, as the "preferred" unit normal ν. The vector field ν orients S by establishing a sense of "left" and "right" in each tangent plane T_p. For example, if $u \in T_p$ and $u \neq 0$ then the cross-product $\nu_p \times u$ is a vector in T_p which is said to point "to the left" of u. If S is the boundary of a domain D whose closure $\bar{D}$ is compact, then we orient S so that ν points away from D.

(d) We denote the element of area on S by dA and the total area of S (if finite) by A.

(B) *Mean Curvature.*

The mean curvature of C^2 surface S is a function $H: S \to \mathbb{R}$ which measures how "bent" S is. We describe H in terms of the more familiar notion of the curvature of a plane curve.

Let P be an oriented plane in $\mathbb{R}^3$. Choose Cartesian coordinates x_1 and x_2 in P so that e_2, the

---

Robert C. Reilly received his M.S. degree (1964) from the Illinois Institute of Technology, where he worked under the direction of Karl Menger, and his Ph.D. (1969) from the University of California, Berkeley, where he wrote his dissertation under the direction of S. S. Chern. He spent 1969–71 at M.I.T. Since 1971 he has been at the University of California, Irvine, except for one year of sabbatical leave at the University of Warwick (England). His primary field of interest is the theory of submanifolds.

unit vector in the positive x_2 direction, points to the left of e_1, the unit vector in the positive x_1 direction. Suppose that Γ is a curve in P which has the parametric representation $(x_1, x_2) = (g_1(s), g_2(s))$, in which g_1 and g_2 are C^2 functions and s is arclength along Γ. Denote the positive unit tangent, $(g_1'(s), g_2'(s))$, by $T(s)$; denote the angle from e_1 to $T(s)$ by $\theta(s)$. (See Fig. 1; positive angles are measured "counterclockwise," i.e., to the left.)

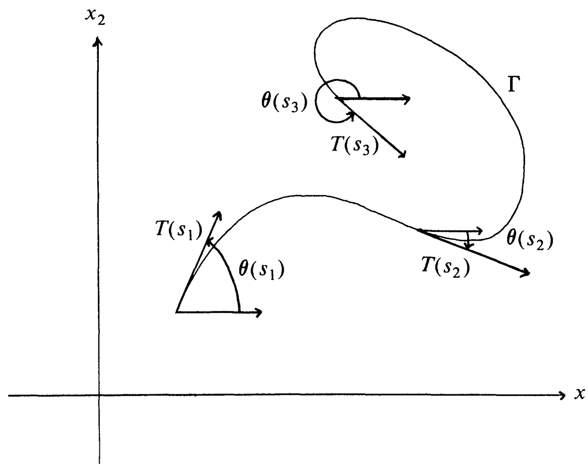


FIG. 1

DEFINITION: The *curvature* of Γ at $(g_1(s), g_2(s))$ is the number $\kappa(s) = d\theta/ds$.

REMARKS.

(a) Suppose that Γ is the graph, $x_2 = \phi(x_1)$, of a C^2 function ϕ and that the positive direction along Γ is in the direction of increasing x_1. Then $\phi' = \tan\theta$, so $\theta = \arctan\phi'$, and $ds = (1+\phi'^2)^{1/2}dx_1$; thus by the chain rule we have

$$\kappa = \frac{d\theta}{ds} = \frac{d\theta}{dx_1}\frac{dx_1}{ds} = \frac{\phi''}{1+\phi'^2}\frac{1}{\left(1+\phi'^2\right)^{1/2}} = \frac{\phi''}{\left(1+\phi'^2\right)^{3/2}}. \tag{1}$$

Note that if $\bar{p} = (\bar{x}_1, \phi(\bar{x}_1))$ is a point of Γ at which the normal line $l_{\bar{p}}$ is parallel to the x_2-axis, then $\phi'(\bar{x}_1) = 0$; in this case (1) takes a simpler form:

$$\kappa(\bar{p}) = \phi''(\bar{x}_1). \tag{1'}$$

Also, note that if κ is positive at a point $p = (x_1, \phi(x_1))$ of Γ then, near p, Γ bends toward the unit vector

$$n(p) = \left(\frac{-\phi'(x_1)}{\left(1+\phi'(x_1)^2\right)^{1/2}}, \frac{1}{\left(1+\phi'(x_1)^2\right)^{1/2}}\right). \tag{2}$$

(This vector is normal to Γ in P.) Likewise, if $\kappa(p) < 0$ then, near p, Γ bends away from $n(p)$.

(b) Suppose that neither the orientation of P nor the positive direction along Γ is given; then at points where $|\kappa| \neq 0$, the sign of κ is ambiguous. (Indeed, in this case there is no well defined notion in P of "counterclockwise rotation.") A convenient way to specify a sign for the curvature at such a point p of Γ is to single out, from the two unit vectors in P normal to Γ at p, a "preferred" unit normal $n(p)$; we say that the curvature at p is positive (with respect to $n(p)$) if Γ bends toward $n(p)$ near p and that it is negative otherwise.

EXAMPLES.

(1) Γ is part of a straight line if and only if $\kappa \equiv 0$.

(2) Γ is part of a circle of radius $R > 0$ if and only if $|\kappa| \equiv 1/R$. More precisely, if $n(p)$ is the outward pointing normal, then $\kappa(p) = -1/R$.

Now suppose that S is an oriented C^2 surface and that p is a point of S. Choose a unit vector e_0 in the tangent plane T_p and for each θ in $[0, 2\pi]$ let e_θ be the vector obtained by rotating e_0 counterclockwise in T_p through an angle θ. The vector e_θ and the normal line l_p span a plane P_θ which intersects S orthogonally at p. Let $\Gamma_\theta = P_\theta \cap S$ be the curve of intersection; denote the curvature of Γ_θ at p, relative to the unit normal ν_p, by $\kappa(\theta, p)$. (This makes sense because ν_p is not only normal to S in $\mathbb{R}^3$, it is also normal to Γ_θ in P_θ. See Fig. 2.)

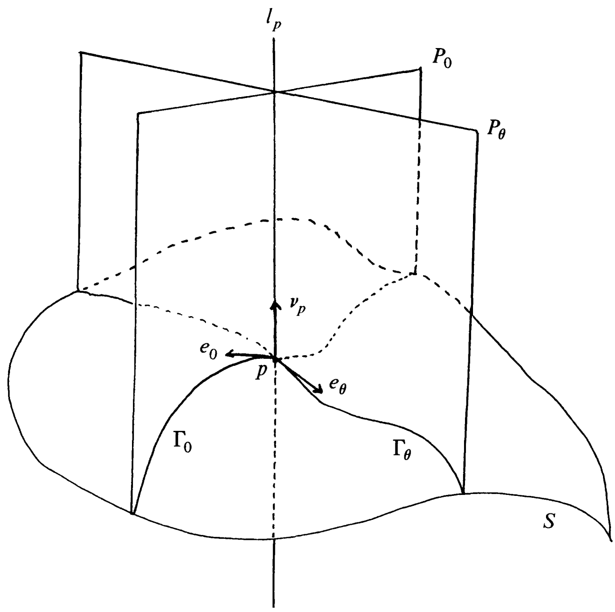


FIG. 2

DEFINITION. The *mean curvature* $H(p)$ of S at p is the average (i.e., the "mean") of the curvatures $\kappa(\theta, p)$ for $\theta \in [0, 2\pi]$. That is,

$$H(p) = \frac{1}{2\pi} \int_0^{2\pi} \kappa(\theta, p)\, d\theta. \tag{3}$$

It is not necessary to calculate $\kappa(\theta, p)$ for all $\theta \in [0, 2\pi]$ in order to determine $H(p)$.

LEMMA. *If $p \in S$ and $\theta \in [0, 2\pi]$, then*

$$H(p) = \frac{1}{2}(\kappa(\theta, p) + \kappa(\theta + \pi/2, p)). \tag{4}$$

Proof. Choose Cartesian coordinates (x_1, x_2, x_3) so that p equals $(0,0,0)$ and l_p is the x_3-axis. Then near p we can express S as the graph of a C^2 function $f(x_1, x_2)$ such that $f_{x_1}(0,0) = f_{x_2}(0,0) = 0$. Let e_0 be the unit vector along the positive x_1-axis and for each θ in $[0, 2\pi]$ construct P_θ and Γ_θ as before. Then (near p) Γ_θ is the graph $x_3 = \phi_\theta(s)$ in the plane P_θ, where the s-axis in P_θ is the intersection of P_θ with the x_1x_2-plane and $\phi_\theta(s) = f(s \cos\theta, s \sin\theta)$. It is clear that $\phi_\theta'(0) = 0$ (because $f_{x_1}(0,0) = f_{x_2}(0,0) = 0$), so we can use equation (1′) to conclude that $\kappa(\theta, p) = \phi_\theta''(0)$.

That is,

$$\kappa(\theta, p) = (\cos^2\theta) f_{x_1x_1}(0,0) + 2(\cos\theta\sin\theta) f_{x_1x_2}(0,0) + (\sin^2\theta) f_{x_2x_2}(0,0). \tag{5}$$

It follows by simple trigonometry that $\kappa(\theta, p) + \kappa(\theta + \pi/2, p) = f_{x_1x_1}(0,0) + f_{x_2x_2}(0,0)$ for all θ in $[0, 2\pi]$; equation (4) is an immediate consequence of this last equation.

3. Examples. On most surfaces the mean curvature varies from point to point, but in this paper we consider only surfaces on which H is constant.

(a) If S is a plane, then each curve Γ_θ is a straight line, so $\kappa(\theta, p) = 0$ for each $p \in S$ and each $\theta \in [0, 2\pi]$; thus $H \equiv 0$.

(b) If S is a sphere of radius R, then each Γ_θ is a circle of radius R, so $\kappa(\theta, p) = -1/R$ for each p and θ; thus $H \equiv -1/R$.

(c) Suppose that S is a right circular cylinder, of radius R, whose axis is a line L. For $p \in S$ let e_0 be a unit vector in T_p parallel to L; then Γ_0 is a straight line (namely, the "generator" through p) and $\Gamma_{\pi/2}$ is a circle of radius R. It follows that $H(p) = \frac{1}{2}(0 + (-1/R)) = -1/2R$.

(d) Let S be the surface formed by rotating a catenary,

$$x_2 = b \cdot \cosh\left(\frac{x_1}{b}\right) \qquad (b = \text{positive constant}),$$

about the x_1-axis. (Such a surface is called a catenoid.) With a little work one can prove that $H \equiv 0$ on S.

The preceding examples possess a great deal of symmetry; those which follow need not possess any.

(e) Carefully dip a loop of wire into a soap solution; a soap film will form, spanning the loop. The stability of such a film requires that its area be smaller than that of any nearby surface spanning the same loop. One can prove (using the Calculus of Variations) that this "minimum area" property implies that the mean curvature must be zero at each point of the film. (Conversely, any surface on which $H \equiv 0$ is called a minimal surface).

(f) Imagine that the loop of wire in the preceding example is the rim of an irregularly shaped container (e.g., the rim of a soap-bubble pipe). Now pump some air into (or out of) this container; the change of air pressure on one side of the soap film spanning the rim causes the film to change its shape. It is proved in the Calculus of Variations that the mean curvature of the deformed soap film is a nonzero constant. If a sufficient amount of air is pumped into the container, then the soap film may escape the rim and form a freely floating soap bubble.

In all but one of the preceding examples, either the surface extends to infinity or it has a boundary curve. The exception is the sphere. The main result in this paper is Aleksandrov's Theorem, which states (in effect) that there is no other exception. More precisely, if the C^4 surface S is the boundary of a compact domain $\bar{D}$ in $\mathbb{R}^3$ and if H is constant on S, then S is a sphere.

Remark. Aleksandrov paraphrases this as "Soap bubbles are round"; see example (f). His proof (see [**1**]) combines a powerful "maximum principle" for nonlinear elliptic equations with a clever geometric argument. In contrast, our proof (which simplifies that given in [**7**]) uses only advanced calculus and, at one point, an existence theorem for a particular boundary-value problem. (It should be noted, however, that Aleksandrov's technique works in a much broader class of problems.) Our proof is in Section 5.

4. Mean Curvature and the Laplacian.

Notation. We denote partial derivatives by subscripts:

$$\frac{\partial F}{\partial x_j} = F_j, \quad \frac{\partial^2 F}{\partial x_j \partial x_k} = F_{kj}, \quad \text{etc.}$$

We denote the first and second directional derivatives along a vector $v = (v_1, \dots, v_n)$ in $\mathbb{R}^n$ by D_v and D_v^2, respectively:

$$D_v F(p) = \frac{d}{dt} F(p + tv)\,|_{t=0} = \sum_{j=1}^{n} F_j(p) v_j,$$

and

$$D_v^2 F(p) = \frac{d^2}{dt^2} F(p + tv)\,|_{t=0} = \sum_{j,k=1}^{n} F_{jk}(p) v_j v_k.$$

DEFINITION. Suppose that F is a C^2 function on an open subset of $\mathbb{R}^n$. The *Laplacian* of F is the function $\Delta F = \sum_{j=1}^{n} F_{jj}$.

REMARK. The value of ΔF is independent of the choice of Cartesian coordinates in $\mathbb{R}^n$.

There is a simple relation between the Laplacian in $\mathbb{R}^2$ and the curvature of plane curves.

THEOREM 1. *Suppose that F is a C^2 function on an open set U in $\mathbb{R}^2$. Let p be a point of U and let Γ be a curve in U passing through p. Denote the restriction of F to Γ by f. Let n_p be a unit normal to Γ at p and let $\kappa(p)$ be the corresponding curvature. Finally, denote arclength along Γ by s. Then*

$$(\Delta F)(p) = \frac{d^2 f}{ds^2}(p) - \kappa(p) D_{n_p} F(p) + D_{n_p}^2 F(p). \tag{6}$$

Proof. Choose Cartesian coordinates x_1, x_2 in $\mathbb{R}^2$ so that $p = (0,0)$ and n_p points in the positive x_2-direction. Then near p the curve Γ is the graph, $x_2 = \phi(x_1)$, of some C^2 function ϕ such that

$$\text{(a)} \quad \phi'(0) = 0 \quad \text{and} \quad \text{(b)} \quad \phi''(0) = \kappa(p). \tag{7}$$

(Part (b) of (7) follows from part (a) and equation (1′).) The Chain Rule implies that

$$\frac{d^2 f}{ds^2} = F_{11} \cdot \left(\frac{dx_1}{ds}\right)^2 + 2F_{12} \cdot \frac{dx_1}{ds}\frac{dx_2}{ds} + F_{22} \cdot \left(\frac{dx_2}{ds}\right)^2$$
$$+ F_1 \cdot \frac{d^2 x_1}{ds^2} + F_2 \cdot \frac{d^2 x_2}{ds^2}. \tag{8}$$

But along Γ we have $d/ds = (1 + \phi'^2)^{-\frac{1}{2}} \dfrac{d}{dx_1}$ and $x_2 = \phi(x_1)$, so (7) implies that at p equation (8) assumes the simpler form

$$\frac{d^2 f}{ds^2}(p) = F_{11}(p) + F_2(p)\kappa(p). \tag{8′}$$

Moreover, because of our choice of coordinates it is obvious that $F_2(p) = D_{n_p} F(p)$ and $F_{22}(p) = D_{n_p}^2 F(p)$. Equation (6) follows easily.

REMARK. If Γ is a circle of radius r and center $(0,0)$, then equation (6) is just the classical formula for the Laplacian in polar coordinates:

$$\Delta F = \frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2} + \frac{1}{r}\frac{\partial F}{\partial r} + \frac{\partial^2 F}{\partial r^2}.$$

There is a similar result for the Laplacian in $\mathbb{R}^3$. Its statement involves an analogue of the quantity $d^2 f/ds^2$ appearing in (6).

DEFINITION: Suppose that G is a C^2 function on an open set W in $\mathbb{R}^3$. Let p be a point of W and let S be a C^2 surface in W passing through p. Denote the restriction of G to S by g. Let $\{P_\theta : 0 \leq \theta \leq 2\pi\}$ be a family of planes through p constructed as in Section 2 (so P_θ contains the normal line l_p and makes an angle θ with P_0) and let $\Gamma_\theta = S \cap P_\theta$ be the corresponding family of curves. Finally, denote the restriction of g to Γ_θ by g_θ and denote the derivatives of g_θ with respect to arclength along Γ_θ by dots ($\dot{g}_\theta$, $\ddot{g}_\theta$, etc.). Then the *surface Laplacian* of g at p is

$$(\Delta_S g)(p) = \ddot{g}_0(p) + \ddot{g}_{\pi/2}(p). \tag{9}$$

REMARKS:
(a) The quantity $\Delta_S g(p)$ makes sense for any C^2 function g on S and does not depend on how we extend g to a C^2 function G in a neighborhood of S in $\mathbb{R}^3$.
(b) If S is a plane, then Δ_S is just the ordinary Laplacian in two variables.
(c) In Riemannian geometry the operator Δ_S is defined using only the "intrinsic geometry" of S. In that context Δ_S is called the Laplace-Beltrami operator.

THEOREM 2. *Suppose that G is a C^2 function on an open set W in $\mathbb{R}^3$. Let p be a point of W and let S be an oriented C^2 surface in W passing through p. Denote the restriction of G to S by g. Let ν_p be the unit normal to S at p and let $H(p)$ be the mean curvature. Denote the surface Laplacian by Δ_S. Then*

$$(\Delta G)(p) = (\Delta_S g)(p) - 2H(p) D_{\nu_p} G(p) + D^2_{\nu_p} G(p). \tag{10}$$

Proof. Choose Cartesian coordinates x_1, x_2, x_3 in $\mathbb{R}^3$ so that $p = (0,0,0)$ and ν_p points in the positive x_3-direction. Let P_0 be the $x_1 x_3$-plane. It is obvious that $G_{33}(p) = D^2_{\nu_p} G(p)$. Moreover, equation (8′) implies that

$$G_{11}(p) = \ddot{g}_0(p) - \kappa(0, p) D_{\nu_p} G(p)$$

and

$$G_{22}(p) = \ddot{g}_{\pi/2}(p) - \kappa(\pi/2, p) D_{\nu_p} G(p).$$

Equation (10) now follows easily from equations (9) and (4).

REMARK. Equation (10) makes sense—and remains true—even if S forms the boundary of W; just extend G to a C^k function on a full neighborhood of S.

EXAMPLE: If S is the sphere $x_1^2 + x_2^2 + x_3^2 = r^2$, then (10) is simply the well-known formula for Δ in spherical coordinates:

$$\Delta G = \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 G}{\partial \phi^2} + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial G}{\partial \theta}\right) + \frac{2}{r}\frac{\partial G}{\partial r} + \frac{\partial^2 G}{\partial r^2}.$$

In particular,

$$\Delta_S = \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \phi^2} + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right).$$

PROPOSITION 1. *Suppose that $g: S \to \mathbb{R}$ is a C^2 function on a C^2 surface S in $\mathbb{R}^3$.*

(a) *If g is constant, then $\Delta_S g = 0$.*

(b) *If S is compact, (e.g., if S is the boundary of a compact domain in $\mathbb{R}^3$), then* $\displaystyle\iint_S \Delta_S g \, dA = 0.$

REMARK. We shall refer to these statements as properties (a) and (b) of Δ_S.

Property (a) follows directly from the definition of Δ_S. Property (b) follows (less directly) from the divergence theorem in $\mathbb{R}^3$; the (somewhat lengthy) proof is in the Appendix. The next result illustrates the power of property (b).

PROPOSITION 2. (Minkowski) *Suppose that S is a compact C^2 surface and that $P: S \to \mathbb{R}$ is the "support function," defined by $P(p) = \langle p, \nu_p \rangle$, where $\langle , \rangle$ is the usual inner product in $\mathbb{R}^3$. Then*

$$\iint_S (H \cdot P + 1)\, dA = 0. \tag{11}$$

Proof. Let $G(x_1, x_2, x_3) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2)$. Then $\Delta G(p) = 3$, $D_{\nu_p} G(p) = P(p)$ and $D_{\nu_p}^2 G(p) = 1$ for each p in S. Thus, equation (10) implies that $HP + 1 = \Delta_S(g/2)$. Equation (11) now follows from property (b) of Δ_S.

5. Aleksandrov's "Soap-Bubble" Theorem. Throughout this section we assume more smoothness than is necessary in order to keep the exposition simple.

THEOREM 3 (Aleksandrov). *Suppose that D is a domain in $\mathbb{R}^3$ whose closure $\overline{D}$ is compact and whose boundary is a C^4 surface S. If the mean curvature of S is constant, then S is a sphere (and thus D is a ball.)*

Our proof of Aleksandrov's Theorem uses several results which ought to be familiar to any mathematics major.

PROPOSITION 3 (Newton's Inequality). *Let $A = (a_{ij})$ be a real $n \times n$ matrix. Define the norm $\|A\|$ and trace $\operatorname{tr}(A)$ by $\|A\|^2 = \sum_{i,j=1}^n a_{ij}^2$ and $\operatorname{tr}(A) = \sum_{j=1}^n a_{jj}$. Then*

$$\text{(a)} \quad \|A\|^2 \geqslant \frac{1}{n}(\operatorname{tr}(A))^2; \tag{12}$$

(b) *we have equality in* (12) *if and only if A is proportional to the $n \times n$ identity matrix: $a_{ij} = c \cdot \delta_{ij}$.*

PROPOSITION 4 (Divergence Theorem). *Let $\overline{D}$ be a compact domain in $\mathbb{R}^3$ whose boundary is a C^4 surface S with outward-pointing unit normal ν. If $F: \overline{D} \to \mathbb{R}$ is a function which is of class C^2 on $\overline{D}$, then*

$$\iiint_D \Delta F \, dx_1 dx_2 dx_3 = \iint_S (D_\nu F)\, dA.$$

(*We use the abbreviation $D_\nu F$ for the function $p \mapsto D_{\nu_p} F(p)$, $p \in S$.*)

PROPOSITION 5. *If $\overline{D}$ and S are as in Proposition 4 and if $P: S \to \mathbb{R}$ is the support function (see Proposition 2), then $\iint_S P\, dA = 3V$, where V is the volume of $\overline{D}$.*

Proposition 3 is the Cauchy-Schwarz inequality as applied to the vectors (a_{ij}) and (δ_{ij}) in $\mathbb{R}^{n^2}$. Proposition 4 is a standard result proved in advanced calculus. Proposition 5 follows directly from Proposition 4 by observing that if

$$F(x_1, x_2, x_3) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2),$$

then $D_\nu F = P$ and $\Delta F = 3$.

Our proof uses one more fact which is not elementary at all.

PROPOSITION 6. *Suppose that $\overline{D}$ and S are as in Proposition 4. Then there exists a function $F: \overline{D} \to \mathbb{R}$ which is C^3 on $\overline{D}$ and C^∞ on D and which solves the following boundary value problem: $\Delta F = 1$ on $\overline{D}$, $F|_S = 0$.*

The proof of Proposition 6 (see [**3**, 6.3-4]) is not elementary, but the statement of this result is very easy to comprehend. Note that we use only the abstract existence of F; it is not important whether we know how to find F.

Proof of Aleksandrov's Theorem: Let $F: \bar{D} \to \mathbb{R}$ be the function described in Proposition 6. Thus $\Delta F = 1$ and $F|_S = 0$, so by equation (10) and property (a) of Δ_S, we see that for each $p \in S$,

$$1 = -2H(p)D_{\nu_p}F(p) + D_{\nu_p}^2F(p). \tag{13}$$

If we multiply both sides of (13) by $D_{\nu_p}F(p)$ and integrate over S, we obtain

$$\iint_S D_\nu F\,dA = \iint_S -2H(D_\nu F)^2 dA + \iint_S (D_\nu F)(D_\nu^2 F)\,dA. \tag{14}$$

Let us analyze the terms appearing in (14):

(i) (Left-hand side of (14).) Because $\Delta F \equiv 1$, Proposition 4 tells us that

$$\iint_S D_\nu F\,dA = V. \tag{15}$$

(ii) (Right-hand side of (14), first term.) Because H is constant, Propositions 2 and 5 imply that

$$H = -A\Big/\iint_S P\,dA = -A/3V.$$

Moreover the Cauchy-Schwarz inequality for integrals, the Divergence Theorem, Proposition 5, and the fact that $\Delta F \equiv 1$ combine to imply that

$$\iint_S (D_\nu F)^2 dA \geqslant \left(\iint_S 1 \cdot D_\nu F\,dA\right)^2 \Big/ \iint_S 1^2 \cdot dA = \left(\iiint_{\bar{D}} \Delta F\,dV\right)^2 \Big/ A = V^2/A.$$

It follows that

$$\iint_S -2H(D_\nu F)^2 dA \geqslant \left(\frac{2A}{3V}\right) \cdot \left(\frac{V^2}{A}\right) = \left(\frac{2}{3}\right)V. \tag{16}$$

(iii) (Right-hand side of (14), second term.) Since $F|_S$ is constant, grad F must be parallel to the unit normal $\nu = (\nu_1, \nu_2, \nu_3)$ at each point of S. That is, grad $F = \langle \text{grad } F, \nu\rangle\nu$; in other words, $F_j = (D_\nu F) \cdot \nu_j (j = 1,2,3)$. Since $D_\nu^2 F = \sum_{i,j=1}^3 F_{ij}\nu_i\nu_j$, it follows that

$$D_\nu^2 F \cdot D_\nu F = \sum_{i,j=1}^3 F_{ij}\nu_i\nu_j D_\nu F = \sum_{i,j=1}^3 F_{ij}\nu_i F_j = \sum_{i=1}^3 \frac{\partial}{\partial x_i}\left(\sum_{j=1}^3 \frac{1}{2}F_j^2\right)\nu_i = D_\nu\left(\frac{1}{2}\,|\text{grad } F|^2\right).$$

Thus the Divergence Theorem implies that

$$\iint_S D_\nu^2 F \cdot D_\nu F \cdot dA = \iiint_D \Delta\left(\frac{1}{2}\,|\text{grad} F|^2\right) dV.$$

But

$$\Delta\left(\frac{1}{2}\,|\text{grad } F|^2\right) = \sum_{i=1}^3 \left(\frac{1}{2}\sum_{j=1}^3 F_j^2\right)_{ii} = \sum_{i,j=1}^3 \left(F_{jii}F_j + (F_{ji})^2\right) = \sum_{i,j=1}^3 (F_{ji})^2$$

(because $\sum_{i=1}^3 F_{jii} = \sum_{i=1}^3 F_{iij} = (\Delta F)_j = (\partial/\partial x_j)(1) = 0$). Combining this with Proposition 3 (Newton's inequality), we see that

$$\Delta\left(\frac{1}{2}\,|\text{grad } F|^2\right) = \sum_{i,j=1}^3 (F_{ij})^2 \geqslant \frac{1}{3}\left(\sum_{j=1}^3 F_{jj}\right)^2 = \frac{1}{3}(\Delta F)^2 = \frac{1}{3}.$$

We conclude that

(a)
$$\iint_S D_\nu^2 F \cdot D_\nu F\,dA \geqslant \frac{V}{3}; \tag{17}$$

(b) inequality (17) is an equation if and only if

$$(F_{ij}) = \frac{1}{3}(\delta_{ij}) \text{ on } \bar{D}. \tag{17'}$$

We can summarize (14), (15), (16), and (17) as follows:

$$V = \iint_S D_\nu F\, dA = \iint_S -2H(D_\nu F)^2\, dA + \iint_S (D_\nu F)(D_\nu^2 F)\, dA \geqslant \frac{2V}{3} + \frac{V}{3}. \tag{18}$$

But $2V/3 + V/3 = V$, so the inequality in (18) must actually be an equation; thus, (17) must also be an equation. By (17′), F must satisfy the system

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{1}{3} \cdot \delta_{ij}, \qquad 1 \leqslant i, j \leqslant 3.$$

We can integrate this system by inspection:

$$F(x_1, x_2, x_3) = \frac{1}{6}(x_1^2 + x_2^2 + x_3^2) + a_1 x_1 + a_2 x_2 + a_3 x_3 + b,$$

where a_1, a_2, a_3, and b are constants of integration. By "completing the square" we see that the zero-set of this quadratic, that is, S, is a sphere.

Concluding Remarks.

(a) Everything we have done here extends immediately to higher dimensions.

(b) H. Hopf [**5**] has conjectured that even in the class of immersed compact surfaces, only the spheres are of constant mean curvature. (An immersed surface may have self-intersections; the definition we have been using does not allow this phenomenon.) He proved that the conjecture is correct for surfaces of genus 0 (i.e., for surfaces homeomorphic to the sphere in $\mathbb{R}^3$.) The conjecture remains open for surfaces of genus $g \geqslant 1$, as does the analogous conjecture—even for manifolds homeomorphic to a sphere—in higher dimensions. (Hopf's proof is based on the uniformization theorem for Riemann surfaces, so is very much "two dimensional.")

(c) Some of the calculations in this paper were inspired by similar computations in a note by Weinberger [**8**].

Appendix.

Proof of Proposition 1, *part b*: Our proof of property (b) for surface Laplacians is based on the observation that equation (10) assumes the particularly simple form

$$(\Delta G)(p) = (\Delta_S g)(p) \quad \text{for} \quad p \in S$$

if G is constant along the normal lines of S. (For example, if S is a sphere centered at the origin, then such a function G must be homogeneous of degree 0.) For small $T > 0$ let

$$W_T = \{q \in \mathbb{R}^3 : q = p + t\nu(p) \quad \text{for some} \quad p \in S \quad \text{and some} \quad t \in [0, T]\},$$

and define $G: W_T \to \mathbb{R}^3$ by the rule $G(p + t\nu(p)) = g(p)$ for all $p \in S$, $t \in [0, T]$. The Inverse-Function Theorem tells us that if T is a sufficiently small positive number, then each q in W_T can be written—in one way only—in the form $p + t\nu(p)$, with $p \in S$ and $t \in [0, T]$; thus G is well defined and smooth. It is obvious that $G|_S = g$ and that G is constant along the normal lines of S. Now for each $t \in [0, T]$ let

$$S_t = \{q \in W_T : q = p + t\nu(p) \quad \text{for some} \quad p \in S\}.$$

We call S_t the *parallel surface* of height t (see Fig. 3). Likewise let $g_t = G|_{S_t}$ and dA_t = element of

[Continued on p. 197.]

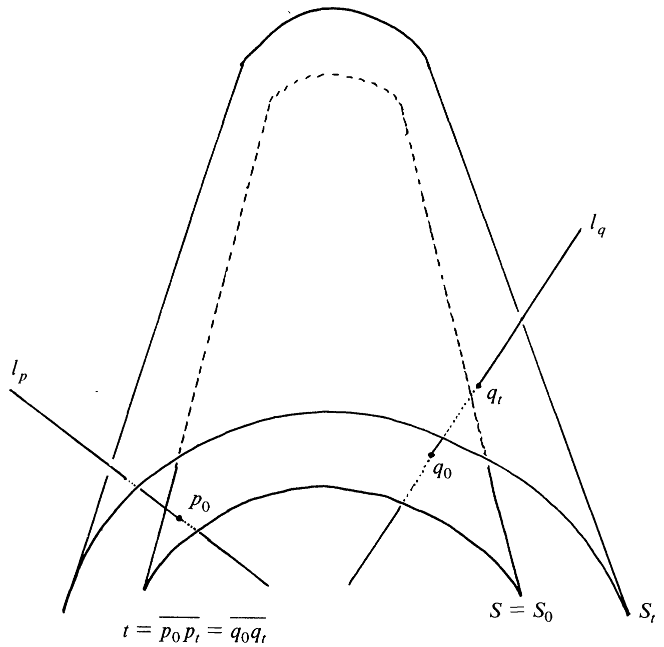


FIG. 3.

area on S_t. It is easy to check that the lines normal to $S = S_0$ are also normal to each of the surfaces S_t for $t \in [0, T]$. Thus equation (10) implies that, for each $t \in [0, T]$ and each $p \in S_t$,

$$\Delta G(p) = \Delta_{S_t} g_t(p). \tag{19}$$

Now integrate ΔG over W_T; from (19) we obtain

$$\iiint\limits_{W_T} \Delta G\, dV = \int_0^T dt \iint\limits_{S_t} \Delta G\, dA_t = \int_0^T dt \iint\limits_{S_t} \Delta_{S_t} g_t\, dA_t. \tag{20}$$

We can use the Divergence Theorem in $\mathbb{R}^3$ to rewrite the leftmost integral in (20) as

$$\iint\limits_{S_T} D_{\nu_T} G\, dA_T - \iint\limits_{S_0} D_{\nu_0} G\, dA_0.$$

However, the normal derivatives $D_{\nu_T} G$ and $D_{\nu_0} G$ vanish because G is constant along the normal lines of S. Thus (20) implies that for each (sufficiently small) $T > 0$,

$$\int_0^T dt \iint\limits_{S_t} \Delta_{S_t} g_t\, dA_t = 0. \tag{21}$$

Property (b) follows easily: differentiate both sides of (21) with respect to T.

References

1. A. D. Aleksandrov, Uniqueness theorems for surfaces in the large V, Vestnik Leningrad Univ. 13 no. 19 (1958) 5–8. (English translation: Amer. Math. Soc. Transl. Ser. 2, 21 (1962) 412–415.)

2. M. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, N.J., 1976.

3. D. Gilbarg and N. S. Trudinger, Elliptic Partial Differential Equations of Second Order, Springer-Verlag, New York, 1977.

4. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.
5. H. Hopf, Über Flächen mit einer Relation zwischen den Hauptkrümmungen, Math. Nachr., 4 (1951) 232–249.
6. R. Millman and G. Parker, Elements of Differential Geometry, Prentice-Hall, Englewood Cliffs, N.J., 1977.
7. R. Reilly, Applications of the Hessian operator in a Riemannian manifold, Indiana Univ. Math. J., 26 (1977) 459–472.
8. H. Weinberger, Remark on the preceding paper of Serrin, Arch. Rational Mech. Anal., 43 (1971) 319–320.

ELEMENTARY QUADRATURES OF ORDINARY DIFFERENTIAL EQUATIONS

LI HONG-XIANG (李鸿祥)
Shanghai Railway Institute, Shanghai 201803, China

Introduction. New types of ordinary differential equations which can be integrated in finite form are of wide interest and appeal to teachers of college mathematics and engineers who apply ordinary differential equations to their work. This MONTHLY has already published many notes on this subject (cf. [**5**]–[**9**], [**11**]). Here we give some new integrable types of first-order and second-order equations; many well-known classical (cf. [**1**]–[**4**]) and modern (cf. [**5**]–[**11**]) integrable types and their integrals are special cases of the results in this paper. We have introduced the concepts of the resolvent function, the characteristic constant and equation, and the discriminant. By using these concepts, the general solutions of many equations can be obtained by algebraic methods or formulas.

1. First-Order Equations.

THEOREM 1.1. *Suppose that $P, Q, F \in C$, $v \in C^1$, $v(x) \neq 0$ in some interval I. If the equality*

$$v' - P(x)v = kQ(x)v^2 \tag{1.1}$$

holds in I, where k is a constant, then the equation

$$y' + P(x)y = Q(x)F(yv(x)) \tag{1.2}$$

is integrable.

Proof. The transformation

$$y = \frac{u}{v} \tag{1.3}$$

converts (1.2) into the form

$$u' = v(x)Q(x)[F(u) + ku]$$

in which the variables are separable.

According to (1.1), after letting $k = 0$ and choosing $v = e^{\int P(x)\,dx}$, equation (1.2) becomes

$$y' + P(x)y = Q(x)F\Big(y\exp\Big(\int P(x)\,dx\Big)\Big), \tag{1.4}$$

and (1.3) becomes

$$y = ue^{-\int P(x)\,dx}. \tag{1.5}$$

It is obvious that, when $P(x) \equiv 0$, the equation (1.4) becomes

$$y' = Q(x)F(y)$$

CENTER SECTION

(Vol. 89, No. 3, Mar. 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S(15-18), L*. Seven Years of Manifold, 1968-1980. Ed: Ian Stewart, John Jaworski. Shiva Pub, 1981, ii + 94 pp, (P). [ISBN: 0-906812-07-0] A marvelous selection from the robust, witty journal Manifold published at Warwick in seven volumes ("years") between 1968 and 1980. The journal vanished with issue #20, leaving behind only its smile: "15 new ways to catch a lion" (including the "kittygory" method based on a forgetful functor); "lemmawocky," an exegesis of Lewis Carroll's poem revealing hidden insights into catastrophe theory; "the multicolored theorem shop;" instructions for knitting a Klein bottle; and much more in the same spirit. Seven Years, like Manifold, also includes some excellent discussion of serious mathematics--simple groups, topology for scientists--and numerous subtle problems encased in delightful, verbal flights of fancy. LAS

General, S*, L. Science Fiction Puzzle Tales. Martin Gardner. Clarkson N. Potter Pub, 1981, xii + 148 pp, $4.95 (P). [ISBN: 0-517-543818] 36 puzzles, primarily mathematical, couched in science fiction settings, reprinted from a column in Isaac Asimov's Science Fiction Magazine. A section of "First Answers" provides resolutions to the problems, as well as new problems arising out of these answers. These problems are resolved in subsequent sections on "Second Answers" and "Third Answers," which also provide historical notes and references. LAS

General, P. Transactions of the Moscow Mathematical Society, 1981, Issue 1. AMS, 1981, v + 250 pp, $91 (P). Translation of Volume 39, 1979. LAS

General, P. Engineering and High Technology Manpower Shortages: The Connection with Mathematics. Eva C. Galambos. Southern Reg Educ Board, 1980, iv + 16 pp, $2 (P). A documented analysis of supply vs. demand for the engineering, computing, physical, and mathematical sciences. Argues that a shortage in these areas will exist at least through the eighties, that this shortage will impede economic growth (both nationally, and in the South), and that the key to recovery is renewed emphasis on a rigorous four-year high school mathematics sequence. LAS

General, P. 18th Scandinavian Congress of Mathematicians: Proceedings, 1980. Ed: Erik Balslev. Prog. in Math., V. 11. Birkhäuser Boston, 1981, ix + 515 pp, $26. [ISBN: 3-7643-3040-6] Eight survey lectures and 24 specialized lectures from the August 1980 Congress in Aarhus, Denmark. LAS

Elementary, T(13: 1), S. Business Mathematics, Second Edition. Esther H. Highland. Reston Pub, 1981, xii + 484 pp, $15.95. [ISBN: 0-8359-0585-3] The approach and emphasis remain unchanged, but new illustrations demonstrate usage of tables and the statistics chapter now includes the calculation and meaning of standard deviation. (First Edition, TR, August-September 1979.) JNC

Mathematics Appreciation, T(13-16: 1), S*, L. Great Moments in Mathematics (Before 1650). Howard Eves. Dolciani Math. Expos., No. 5. MAA, 1980, xiv + 270 pp, $22. [ISBN: 0-88385-305-1] 20 historically-rooted lectures, each followed by numerous exercises, selected and abridged ("cruelly condensed") from the author's popular mathematics appreciation course. Each lecture uses a mixture of biographical anecdote and informal discussion of original sources to introduce vignettes of ancient mathematics with lasting value. Concludes with a section of hints for the exercises. LAS

Mathematics Appreciation, T(13). Modern Mathematics: An Elementary Approach, Alternate Edition. Ruric E. Wheeler. Brooks/Cole, 1981, xvii + 442 pp, $18.95. [ISBN: 0-8185-0413-7] A flexible introductory or remedial text for liberal arts mathematics. Emphasizes problem-solving, developing numerical and geometric intuition. Introduces number systems, set theory, logic, geometry, number theory, probability and statistics, financial terms, estimation (including calculator error). Unusually many and varied exercises, both routine and discovery-oriented. PZ

Mathematics Appreciation, S*(13-15), P, L.** The Role of Mathematics in the Rise of Science. Salomon Bochner. Princeton U Pr, 1981, x + 386 pp, $6.95 (P). [ISBN: 0-691-02371-9] Paperback edition of a classic first published in 1966 (TR, May 1967; ER, October 1967). Essentially a collection of

separate essays (each previously published) on the Greek roots of mathematics, on the relation of mathematics to physics, and on the role of symbols in myth and mathematics. Concludes with brief idiosyncratic biographical sketches of all important persons named in the essays. LAS

Precalculus, T(13: 1). Fundamental College Algebra, Second Edition. Mervin L. Keedy, Marvin L. Bittinger. Addison-Wesley, 1981, xi + 498 pp, $15.95. [ISBN: 0-201-03847-1] This edition includes additional topics, additional exercises with a greater level of conceptual difficulty, and exercises designed for calculator-aided solution. Uses transformations to simplify topics related to graphing and quadratic functions. JJ

Precalculus, T(13: 2). Background for Calculus, Second Printing. Ada Peluso. Kendall/Hunt Pub, 1980, 604 pp, $15.95 (P). [ISBN: 0-8403-2295-X] Standard presentation of algebra and trigonometry with an emphasis on functions. Best feature is the great number of examples used to illustrate each new concept or process. (First Edition, TR, April 1979.) JJ

Precalculus, S(13). Sets, Functions and Logic: Basic Concepts of University Mathematics. Keith J. Devlin. Chapman & Hall, 1981, ix + 90 pp, $7.95 (P); $14.95. [ISBN: 0-412-22670-7; 0-412-22660-X] Achieves its goal of introducing "the formal use of mathematics...in a non-formalistic manner." Its conversations gently explore the abstract nature and notation of mathematical proof, as well as the related properties of logical connectives and quantifiers. JJ

Education, S(16-18). How to Evaluate Your Mathematics Program. NCTM, 1981, v + 18 pp, $2.50 (P). [ISBN: 0-87353-183-3] Ninety-three-point check list organized into 21 standards under general headings of instruction, materials, teachers and physical facilities. Useful guide for assessing strengths and weaknesses and setting goals for improvement. MW

Education, P. How to Develop Problem Solving Using a Calculator. Janet Morris. NCTM, 1981, v + 42 pp, $4 (P). [ISBN: 0-87353-175-2] Using Polya's 4-step heuristic as a context and elementary school mathematics as content, the text focuses on three problem-solving techniques: guess-and-check, make a chart/list, and look for a pattern. Throughout, the calculator is used as a helpful tool, not a technique in itself. JJ

Education, P. Math Activities for Child Involvement, Third Edition. C.W. Schminke, Enoch Dumas. Allyn & Bacon, 1981, xi + 336 pp, $11.95 (P). [ISBN: 0-205-07302-6] This edition expands the hand calculator section and provides additional activities for both early childhood education and special education. The index of skills by activity make the text an invaluable resource for any elementary school teacher. JJ

Education, P, L.** Teaching Mathematics: What Is Basic? Stephen S. Willoughby. Council for Basic Education, 1981, 46 pp, $2 (P). A thoughtful rationale for school mathematics by an experienced mathematics educator. Includes a sensible commentary on the new math, a homely analysis of textbook mathematics, and advice to the public on evaluation of curricula, textbooks and teachers. Should be required reading for all school administrators and board members. LAS

Education, P.** Rethinking Mathematical Concepts. Roger F. Wheeler. Halsted Pr, 1981, 314 pp, $55.95. [ISBN: 0-470-27116-7] An innovative, common-sense discussion of priorities and pitfalls facing beginning teachers of calculus. Its purpose is to encourage teachers to think about the presentation of mathematical concepts, especially about choices of definition and notation that have long-range pedagogical implications. Even experienced teachers could learn much from this book, if they could afford its outrageous price. LAS

Education, S(16). Activities for Junior High School and Middle School Mathematics. Kenneth E. Easterday, Loren L. Henry, F. Morgan Simpson. NCTM, 1981, 218 pp, $8.25 (P). [ISBN: 0-87353-189-2] Readers of the Arithmetic Teacher and the Mathematics Teacher will recognize some of their favorite articles here. Activities are not repeated from other NCTM collections. Lots of ideas for teaching junior high-middle school topics, from counting and place value through geometry, logic and problem solving. MW

Education, T(15-17), S, P, L. Motivated Mathematics. A. Evyatar, P. Rosenbloom. Cambridge U Pr, 1981, viii + 284 pp, $39.50 (P). [ISBN: 0-521-23308-9] A handbook of applications and mathematical models for prospective or practicing secondary teachers. The mathematical topics cover most of the precalculus and high school calculus courses; the applications range from Malthus and Volterra population models to Zipf phenomena and cryptography. Much of this material was developed during the 1960's as an alternative to SMSG curricula. While its then-prophetic message has been somewhat diluted by the numerous current textbooks that include these same applications, its value in teacher education remains high. LAS

Education, P*, L. Mathematics Teaching Today: Perspectives from Three National Surveys. James T. Fey. NCTM, 1981, ii + 30 pp, $3 (P). [ISBN: 0-87353-186-8] Reprints of two articles (Arithmetic Teacher, October 1979; Mathematics Teacher, October 1979) reporting on an NCTM project to review and interpret the findings of an NSF study of school science and mathematics education. Extensive quotations from teachers help illustrate both their frustrations and their aspirations in the post-"new math" era. "The most discouraging feature...is the consistent pattern of great differences between apparent reality...and the recommendations...of prominent teachers, supervisors and professional organizations." LAS

History, P, L.** The Origins of Cauchy's Rigorous Calculus. Judith V. Grabiner. MIT Pr, 1981, x + 252 pp, $25. [ISBN: 0-262-07079-0] Cauchy's 1823 Calcul infinitésimal signalled a "true scientific revolution," the beginning of a century's work on the foundations of calculus. Grabiner employs the historian's tools to seek roots for this revolution. She finds them primarily in the works and aspirations of Lagrange--the first major mathematician to take seriously the occasional criticisms of Newtonian and Eulerian rigor. Lagrange's proposal to found calculus on algebra provides a rich context for the work of both Cauchy and Bolzano, and helps explain (without accusations of plagiarism) the many similarities in their works. This volume includes in an appendix translations of key parts of Cauchy's work, as well as extensive notes relating Grabiner's analysis to original sources. LAS

History, P, L*. The Mathematical Work of John Wallis, Second Edition. J.F. Scott. Chelsea Pub, 1981, xi + 240 pp, $12.95. [ISBN: 0-8284-0314-7] A thorough, informative analysis of mid-seventeenth century arithmetic, mechanics and algebra, via explication of the major treatises of Wallis, "Newton's great precursor." Republication of the 1938 London original edition. Includes in an Appendix brief biographies of Wallis's contemporaries mentioned in the volume, and examples of the evolving algebraic notation of the era. LAS

History, P, L. Social History of Nineteenth Century Mathematics. Ed: Herbert Mehrtens, Henk Bos, Ivo Schneider. Birkhäuser Boston, 1981, xii + 301 pp, $24.95. [ISBN: 3-7643-3033-3] Papers from a July 1979 conference in Berlin devoted to mathematics as a cultural activity: a study of historical and social processes influencing the development of mathematics (e.g., professionalization of mathematics, development of schools of mathematics) and of social forces directed in part by the maturing of mathematics as an intellectual and scientific activity. LAS

History, P, L. History of Binary and Other Nondecimal Numeration. Anton Glaser. Tomash Pub, 1981, xiii + 218 pp, $28. [ISBN: 0-938228-00-5] Reprint of 1971 original edition (TR, October 1971), itself based on the author's doctoral thesis. Consists primarily of a chronological sequence of examples from 1600 to 1946 of references to nondecimal number systems, together with numerous tables and a thorough bibliography. The discussion of computers and their relation to nondecimal numeration is now especially out-of-date. LAS

Foundations, P. Analytic Sets. C.A. Rogers, et al. Academic Pr, 1980, x + 499 pp, $115.50. [ISBN: 0-12-593150-6] Five chapters, written for a July 1978 conference at University College, London, covering K-analytic sets, capacity, Souslin schemes, infinite games, effective descriptive set theory, non-separable metric spaces. Concludes with a problem list from the conference. LAS

Foundations, P. In the Labyrinths of Language: A Mathematician's Journey. V.V. Nalimov. ISI Pr, 1981, xix + 246 pp, $22.50. [ISBN: 0-89495-007-X] A provocative exploration of human language (from mathematics at one extreme to art and Hindu metaphysics at another) as a probabilistic and cybernetic enterprise whose objective is understanding in the face of uncertainty. Translated from a 1974 Russian monograph Probabilistic Model of Language. LAS

Foundations, T?(17-18: 2), P. Axiomatic Set Theory: Impredicative Theories of Classes. Rolando Basim Chuaqui. Math. Stud., V. 51. Elsevier North Holland, 1981, xv + 388 pp, $48.75 (P). [ISBN: 0-444-86178-5] Strong first order set theories based on the unrestricted (impredicative) comprehension schema for classes of sets (Bernays, Morse-Kelley-Tarski). Topics chosen to highlight the elegance and power of strong principles of reflection and recursion to the limits of Bernays' theory (no metatheory, partition theory or large large cardinals, but small large cardinals and cardinals without choice). Will seem excessively formal to most readers, but Chuaqui defends his deliberate choice of style. Price and number of typographical errors both seem high. GHM

Foundations, S*(16-17), P, L*. Mathematics and Physics. Yu. I. Manin. Prog. in Physics, V. 3. Transl: Ann and Neal Koblitz. Birkhäuser Boston, 1981, xii + 99 pp, $10. [ISBN: 3-7643-3027-9] A fascinating, subjective reflection on the nature of mathematics and physics, suitable for philosophically-minded audiences with at least a thorough undergraduate training in both subjects. In addition to discussion of standard topics such as observables and space-time, Manin also tours the "gallery of geometric shapes" used in modern physics--catastrophes, solitons, manifolds, gauge theories, Feynman's integrals, et al. LAS

Number Theory, T(18), P, L. The Hardy-Littlewood Method. R.C. Vaughan. Cambridge Tracts in Math., V. 80. Cambridge U Pr, 1981, 11 + 172 pp, $34.50. [ISBN: 0-521-23439-5] An engaging introduction to the H-L method and its application to the Waring and Goldbach problems and related diophantine problems. Contains an extensive bibliography and some exercises. SG

Number Theory, T(16-17), L. Zetafunktionen und quadratische Körper. D.B. Zagier. Springer-Verlag, 1981, ix + 144 pp, $12.80 (P). [ISBN: 0-387-10603-0] An introduction to the algebraic and analytic features of binary quadratic forms. The first half develops properties of Dirichlet series. The second develops form theory (genus, representation, class numbers) and quadratic field theory, culminating in a beautiful theorem relating the class number of $Q(\sqrt{-p})$, $p \equiv 3$ (4), to the continued fraction expansion of $\sqrt{p}$. Many examples and exercises are included. SG

Algebra, S(16-17), P, L. N-dimensional Crystallography. R.L.E. Schwarzenberger. Res. Notes in Math., V. 41. Pitman Pub, 1980, 139 pp, $15.95 (P). [ISBN: 0-8224-8468-4] Somewhat terse and formal treatise on symmetry groups in general n-dimensional Euclidean space. Sparse on illustrations. The approach is geometric but presupposes skill in group theory and knowledge of symmetry. Explicit

enumeration of all 17 plane and 230 space symmetry groups precedes development of qualitative results valid in arbitrary dimensions. GHM

Algebra, P, L. The Theory of Spinors. Élie Cartan. Dover Pub, 1981, 157 pp, $4 (P). [ISBN: 0-486-64070-1] Republication of the 1966 Hermann translation of the original 1937 French edition that gives a geometrical definition of spinors in space of n dimensions. Cartan's classic provides a rigorous basis for the Dirac equation, and helps link Riemannian geometry with quantum mechanics and relativity theory. LAS

Algebra, T, P, L. Methods of Representation Theory with Applications to Finite Groups and Orders, Volume 1. Charles W. Curtis, Irving Reiner. Wiley, 1981, xxi + 819 pp, $55. [ISBN: 0-471-18994-4] How can you describe this monumental treatise in 50 words or less? Neither a sequel to nor a mere revision of the old Curtis-Reiner. Covers ordinary and modular representation, orders and lattices, local and global integral representation theory, plus 200 pages of algebraic preliminaries. A must for every algebraist. SG

Finite Mathematics, T(13: 1, 2). Finite Mathematics with Applications to Business and the Social Sciences. Ruric E. Wheeler, W.D. Peeples, Jr. Brooks/Cole, 1981, xiii + 577 pp, $21.95. [ISBN: 0-8185-0418-8] An introductory text flexible enough to allow an emphasis on business applications, on the social sciences, or on probability and statistics. Problem sets are of graded difficulty and include calculator-dependent exercises. JJ

Calculus, S(13). Basic Mathematics for Biochemists. Athel Cornish-Bowden. Chapman and Hall, 1981, ix + 137 pp, $12.95 (P); $23. [ISBN: 0-412-23010-0; 0-412-23000-3] A good summary of useful mathematics for students in biology and chemistry. Includes exponents and logarithms, differential and integral calculus, and partial differentiation. LLK

Calculus, S?? Meta-Calculus: Differential and Integral. Jane Grossman. Archimedes Foundation, 1981, v + 30 pp, $3 (P). A privately published pamphlet setting forth an eccentric interpretation of measures and weighted integrals. LAS

Complex Analysis, T(17-18). Introduction to Riemann Surfaces, Second Edition. George Springer. Chelsea Pub, 1981, viii + 309 pp, $13.95. [ISBN: 0-8284-0313-9] Corrected reprint of the 1957 Addison-Wesley original edition. A self-contained modern yet elementary introduction. LAS

Complex Analysis, P. Recent Developments in Several Complex Variables. Ed: John E. Fornaess. Annals of Math. Stud., No. 100. Princeton U Pr, 1981, xi + 452 pp, $32.50; $12.50 (P). [ISBN: 0-691-08285-5; 0-691-08281-2] 26 papers from a conference on several complex variables held at Princeton University in April 1979. Most are research reports, some are survey articles. PZ

Differential Equations, S(14-15), L**.** Modelling with Differential Equations. D.N. Burghes, M.S. Borrie. Halsted Pr, 1981, 172 pp, $29.95. [ISBN: 0-470-27101-9] A wonderful source of examples illustrating the use of ordinary differential equation models in the physical, biological, and social sciences. Each chapter presents models which use a particular type of differential equation (separable, linear first order, etc.). AO

Differential Equations, P. Topics in Dynamic Bifurcation Theory. Jack K. Hale. Reg. Conf. Ser. in Math., No. 47. AMS, 1981, iii + 84 pp, $6.40 (P). [ISBN: 0-8218-1698-5] Proceedings of a CBMS conference held at Texas-Arlington in 1980. Among the topics discussed are first order bifurcation, two dimensional periodic systems, higher order bifurcation near equilibrium, and bifurcation in infinite dimensions. SG

Differential Equations, S(17), P. Solvability of Nonlinear Equations and Boundary Value Problems. Svatopluk Fučík. D Reidel Pub, 1980, 390 pp, $29.95. [ISBN: 90-277-1077-5] An account of the solvability of equations of the form Lu = Su in Banach spaces, where L is a linear and S is a non-linear mapping. Includes recent results and has an extensive bibliography. JG

Differential Equations, T(17-18), S, L. Plane Waves and Spherical Means Applied to Partial Differential Equations. Fritz John. Springer-Verlag, viii + 172 pp, $18 (P). [ISBN: 0-387-90565-0] Paperback republication of the 1955 Interscience classic tract that introduced plane and spherical integrals as a tool in partial differential equations. LAS

Differential Equations, P. Three Papers on Dynamical Systems. A.G. Kusnirenko, A.B. Katok, V.M. Alekseev. AMS Transl., Ser. 2, V. 116. AMS, 1981, v + 169 pp, $32.40. [ISBN: 0-8218-3066-X] Translations of the survey lectures from the 1971 Mathematics Summer School at Kaciveli, U.S.S.R., emphasizing contemporary results in the general problems of classification of dynamical systems, and applications to qualitative results in celestial mechanics. LAS

Numerical Analysis, P. Sequence Transformations and Their Applications. Jet Wimp. Math. in Sci. and Eng., V. 154. Academic Pr, 1981, xix + 257 pp, $38.50. [ISBN: 0-12-757940-0] This monograph discusses practical methods for accelerating the convergence of sequences. The techniques covered are applicable to a wide variety of problems in numerical analysis which are based on the evaluation of the limit of a sequence. AO

Numerical Analysis, T*(15-17: 1), S, L. Matrix Computations and Mathematical Software. John R. Rice. McGraw-Hill, 1981, xii + 248 pp, $23.95. [ISBN: 0-07-052145-X] This textbook discusses

numerical linear algebra (the solution of linear systems and least squares problems) and the design and evaluation of mathematical software for matrix computations. The final chapter contains eighteen projects suitable for small groups of students. An interesting book. AO

Numerical Analysis, S(15-17), P*. Computational Techniques for Ordinary Differential Equations. Ed: I. Gladwell, D.K. Sayers. Academic Pr, 1980, xi + 303 pp, $26. [ISBN: 0-12-285780-1] This volume contains the papers presented by the invited speakers at a conference held at the University of Manchester in December 1978. Both initial value and boundary value problems are discussed. The expository papers included may be of interest to those who are not experts in this field. AO

Analysis, P, L. Topics in Iteration Theory. György Targonski. Vandenhoeck & Ruprecht, 1981, 292 pp, DM 45.00 (P). [ISBN: 3-525-40146-9] A well-written survey of some recent developments. Topics include: orbits, iterative roots, continuous and analytic iteration, entropy, chaos, and applications to ecology and turbulence. SG

Analysis, P. Abstract Cauchy Problems and Functional Differential Equations. F. Kappel, W. Schappacher. Res. Notes in Math., V. 48. Pitman, 1981, 238 pp, $22.95 (P). [ISBN: 0-273-08494-1] Proceedings of a June 1979 workshop on functional differential equations and nonlinear semigroups held at the Volksbildungsheim Schloss Retzhof. LAS

Analysis, P. Nonlinear Problems of Analysis in Geometry and Mechanics. M. Atteia, D. Bancel, I. Gumowski. Res. Notes in Math., V. 46. Pitman, 1981, 208 pp, $21.95 (P). [ISBN: 0-273-08493-3] Proceedings of an April 1979 symposium held at Paul-Sabatier University in Toulouse, France. LAS

Differential Geometry, S, P*, L. Singularity Theory: Selected Papers. V.I. Arnold. London Math. Soc. Lect. Note Ser., V. 53. Cambridge U Pr, 1981, 266 pp, $27.50 (P). [ISBN: 0-521-28511-9] Reprint from Russian Mathematical Surveys of seven seminal papers published between 1968 and 1978 on the classification of critical points of smooth functions. A useful introduction to Arnold's work, "one of the most beautiful discoveries in mathematics in recent years," which subsumes and extends Thom's analysis of the elementary catastrophes. LAS

Geometry, P. Geometry--von Staudt's Point of View. Ed: Peter Plaumann, Karl Strambach. D Reidel Pub, 1981, xi + 430 pp, $58. [ISBN: 90-277-1283-2] A photo-offset edition of the proceedings of the NATO Advanced Study Institute held in Bad Windsheim, West Germany, in July 1980 which attempted to correct the past neglect of von Staudt's thesis (contained in his theorem known as the Fundamental Theorem of Projective Geometry). JNC

Geometry, S(14), P, L.** Higher Geometry. N.V. Efimov. Trans: P.C. Sinha. MIR Pub, 1980, 560 pp, $11. A thorough, lucid presentation of the fundamentals of Euclidean, non-Euclidean, and projective geometry and the geometrical aspects of special relativity. Translated from the sixth Russian edition (1978); an excellent and reasonably priced reference. No exercises. JNC

Geometry, T(17), S, P, L. Affine Planes with Transitive Collineation Groups. Michael J. Kallaher. Elsevier North Holland, 1982, xiii + 155 pp, $34.95. [ISBN: 0-444-00620-6] A book suitable for graduate students and researchers in the field of combinatorics and finite geometries. The first 8 chapters are an introduction to the main concepts of projective and affine planes. Later chapters include Andre's Theorem, Finite Affine Line Transitive Collineation Groups, Block Orbits. Several unsolved problems are stated. JG

Operations Research, T(15-16: 1), S*, L*. Foundations of Analysis in Operations Research. J. William Schmidt, Robert P. Davis. Oper. Res. and Industrial Eng. Academic Pr, 1981, xi + 383 pp, $27. [ISBN: 0-12-626850-9] Designed for students with an undergraduate background in linear algebra and multivariable calculus, this text provides the mathematical foundation for a first course in operations research methodology. AO

Optimization, T(15-17: 1, 2), S, L. Combinatorial Optimization: Algorithms and Complexity. Christos H. Papadimitriou, Kenneth Steiglitz. Prentice-Hall, 1982, xvi + 496 pp, $34. [ISBN: 0-13-152462-3] A tour of contemporary algorithms joining mathematical programming with computational complexity: five chapters on linear programming are followed by six on flow and matching problems; the Soviet ellipsoid algorithm provides a "missing link" to NP problems. Concluding chapters treat approximation algorithms, branch and bound methods, and local search techniques. Algorithms are expressed in "pidgin Algol." LAS

Probability, P. Point Processes and Queuing Problems. Ed: P. Bártfai, J. Tomkó. North-Holland Pub, 1981, 426 pp, $63.50. [ISBN: 0-444-85432-0] Proceedings of a colloquium sponsored by the János Bolyai Mathematical Society at the University of Debrecen (Hungary) in September, 1978. Contains 24 papers by specialists from 18 different countries. LAS

Computer Literacy, S(13). Computer Consciousness: Surviving the Automated 80s. H. Dominic Covvey, Neil Harding McAlister. Addison-Wesley, 1980, xi + 212 pp, $5.95 (P). [ISBN: 0-201-01939-6] Layman's overview of the fundamentals and terminology of computer technology. Primary weakness is that most of its references and data are already out-of-date. JJ

Computer Programming, S. Some Common Basic Programs, TRS-80 Level II Edition. Lon Poole, Mary Borchers, Karl Koessel. Osborne/McGraw-Hill, 1981, x + 193 pp, $14.99 (P). [ISBN: 0-931988-54-3] Not a text. 76 programs written to be keyed into a TRS-80. Also available on cassettes. LLK

Computer Programming, T(13: 1), S. Programming with FORTRAN 77. J. Ashcroft, et al. Granada Pub, 1981, 294 pp, (P). [ISBN: 0-246-11573-4] A good readable text. Examples and exercises at the right level. LLK

Software Systems, P. On the Construction of Programs. Ed: R.M. McKeag, A.M. Macnaghten. Cambridge U Pr, 1980, x + 422 pp, $24.50. [ISBN: 0-521-23090-X] This volume contains the texts of papers contributed by the speakers at a summer school held in Belfast in 1979. The first half describes the application of structured programming techniques to large-scale programs. The second half describes current research in programming, with a particular emphasis on parallelism. AO

Computer Science, S. Guide to Systems Applications. John P. Grillo, J.D. Robertson. Wm C Brown, 1981, xi + 268 pp, $17.95 (P) [ISBN: 0-697-09952-0]; Data Management Techniques. 1981, x + 193 pp, $16.95 (P) [ISBN: 0-697-09954-7]; Introduction to Graphics. 1981, viii + 133 pp, $15.95 (P) [ISBN: 0-697-09953-9]; Techniques of Basic. 1981, xv + 256 pp, $18.95 (P). [ISBN: 0-697-09951-2] This is a Personal Computer Series designed specifically for the TRS-80 but adaptable to microcomputers in general. Each contains a large number of tested programs which can be keyed into the TRS-80. LLK

Computer Science, T(13: 1), S. An Introduction to Mini & Micro Computers. Fabian Monds, Robert McLaughlin. Peter Peregrinus, 1981, x + 133 pp, $27.50 (P). [ISBN: 0-906048-48-6] Pulls together many details of hardware and software for someone with little background experience. However, non-standard symbols are used for logic gates and timing diagrams and some topics are out of date (i.e., core arrays aren't even used now). LLK

Control Theory, P. Complex Variable Methods for Linear Multivariable Feedback Systems. Ed: A.G.J. MacFarlane. Taylor & Francis, 1980, 368 pp, $39.95. [ISBN: 0-85066-197-8] Reprints of 14 papers published between 1976 and 1979 in the International Journal of Control, supplemented by brief section synopses, a common bibliography and a (short) unified index. The theme is to extend the classical theory of single-loop feedback systems to multivariable cases; the major tool is analytic function theory, used to study the behavior (zeros, poles, root-locus asymptotes) of characteristic gain and frequency functions associated with transfer function matrices. LAS

Systems Theory, T(18: 1, 2), P. Linear Systems and Operators in Hilbert Space. Paul A. Fuhrmann. McGraw-Hill, 1981, x + 325 pp, $44.95. [ISBN: 0-07-022589-3] Using results of modern algebra and operator theory which are developed in the text, the author extends to Hilbert space elements of the theory of finite-dimensional linear systems. The book is intended for the use of both engineers and mathematicians, and could be used as a text at the graduate level. PZ

Systems Theory, P. Cybernetics in Biology and Medicine, Systems Analysis, Systems Engineering Methodology, Mathematical Systems Theory. Ed: Franz R. Pichler, Robert Trappl. Prog. in Cybernetics and Systems Res., V. VI. McGraw-Hill, 1982, xiii + 398 pp, $80. [ISBN: 0-07-049846-6] First half of proceedings of the 1978 European Meeting on Cybernetics and Systems Research, held in Linz, Austria. (Second half will appear as Volume VII.) LAS

Applications (Archaeology), S(15-17), P, L. Mathematics in Archaeology. Clive Orton. Humanities Pr, 1980, 248 pp, $33.75. [ISBN: 0-00-216226-1] A well-written, formula-free introduction to numerous mathematical techniques used in archaeological research (describing shapes, clustering techniques, stratigraphy, seriations, radiocarbon calibration, discriminant analysis, mapping techniques, trend surface analysis, simulation, dimensional analysis, quantification from sherds and bones) arranged under such homely headings as "What is it?", "How old is it?", "Where does it come from?", and illustrated with numerous examples mostly from English sites. A final chapter suggests possible future trends, e.g., computer methods and catastrophe theory. Provides considerable insight into the ways in which archaeologists can draw conclusions from incomplete and fragmentary data. A good reference for introductory courses in data analysis. LAS

Applications (Biology), P. Mathematical Biology: A Conference on Theoretical Aspects of Molecular Science. Ed: T.A. Burton. Pergamon Pr, 1981, vi + 249 pp, $30. [ISBN: 0-08-026348-8] Proceedings of a May, 1980 conference at Southern Illinois University in Carbondale. 14 substantial papers on diverse topics, each serving to introduce both central themes and current frontiers. LAS

Applications (Biology), P. Mathematical Aspects of Physiology. Ed: Frank C. Hoppenstedt. Lect. in Appl. Math., V. 19. AMS, 1981, vi + 394 pp, $38. [ISBN: 0-8218-1119-3] Proceedings of the 1980 AMS-SIAM summer seminar held at the University of Utah. Features a 100 page survey article by C.S. Peskin, as well as several in-depth case studies and background lectures. Most papers concern blood circulation or neural activity. LAS

Applications (Economics), T(15-16: 1), P, L. Mathematical Methods in Finance and Economics. Sarkis J. Khoury, Torrence D. Parsons. Elsevier North Holland, 1981, xiii + 295 pp, $39.50. [ISBN: 0-444-00425-4] Treats mathematical aspects of economic models, value calculations, Monte Carlo methods for investment decisions. Matrix algebra, with tableaus and pivoting, is developed to allow discussion of linear and nonlinear optimization, Lagrange and Kuhn-Tucker conditions for constrained optimization, game theory. Helpful examples are given throughout. PZ

Applications (Economics), P. Lecture Notes in Economics and Mathematical Systems-188: Economic Theory of Public Enterprise. Dieter Bös. Springer-Verlag, 1981, vii + 142 pp, $12 (P). [ISBN: 0-387-10567-0] "A book about prices" of publicly supplied goods. Assuming that public enterprise seeks to maximize welfare (rather than profit), Bös shows that marginal cost pricing is best, until

deficits impose political constraints. The major part of this book concerns models for second-best pricing policies based on a mixture of political and economic constraints. LAS

Applications (Economics), S, P*, L.** Handbook of Mathematical Economics, Volume 1. Ed: Kenneth J. Arrow, Michael D. Intriligator. Handbooks in Econ., Book 1. Elsevier North Holland, 1981, xvii + 378 pp, $150. [ISBN: 0-444-86126-2] First of a three volume state-of-the-art survey of mathematical (not statistical) economics. This volume treats topology, convexity, mathematical programming, global analysis, dynamical systems, and control, measure, probability, and game theory, and applies each topic to economic theory. Opens with an historical essay on the development of mathematical economics, and concludes with a comprehensive index and list of theorems. A magnificant resource for both economists and mathematicians that puts surprising, deep mathematics to significant and often unexpected use. LAS

Applications (Engineering), P, L. Image Reconstruction from Projections: The Fundamentals of Computerized Tomography. Gabor T. Herman. Comp. Sci. and Appl. Math. Academic Pr, 1980, xiv + 316 pp, $29.50. [ISBN: 0-12-342050-4] A comprehensive nuts-and-bolts survey of the physics, the mathematics, and the computer methods which make possible such diverse scientific products as CAT body scanners and calculation of the solar corona. Descriptions of various reconstruction algorithms (Radon transform methods, series expansion methods) compose the major part of the book. Proofs and derivations are contained in a mathematical appendix. LAS

Applications (Engineering), P. Power from Sea Waves. Ed: B. Count. Inst. of Math. and its Appl. Academic Pr, 1980, xv + 449 pp, $57. [ISBN: 0-12-193550-7] Papers from a June 1979 conference in Edinburgh, Scotland: global energy scene, measurement and prediction of ocean waves, and analysis of energy devices operating in waves. LAS

Applications (Environmental Science), P. Air Pollution Modeling and Its Application I. Ed: C. De Wispelaere. NATO: Challenges of Mod. Soc., V. 1. Plenum Pr, 1981, xiii + 747 pp, $75. [ISBN: 0-306-40820-1] Proceedings of a November 1980 NATO technical meeting in Amsterdam on interregional transport of air pollutants. A rich source of important contemporary mathematical models. LAS

Applications (Physics), L. A Dictionary of Scientific Units Including Dimensionless Numbers and Scales, Fourth Edition. H.G. Jerrard, D.B. McNeill. Chapman & Hall, 1980, ix + 212 pp, $16.95. [ISBN: 0-412-22360-0] Definitions for about 800 named units and dimensionless numbers, from archaic ("cé" for one-hundredth of a day; to contemporary ("byte"), from sublime ("cubit") to ridiculous ("Skewes number" for 10^{3800}, "an indication of the number of primes"). Appendices provide tables of weights and measures, of fundamental physical constants, and of conversion factors for SI and CGS units. LAS

Applications (Physics), S*(15-17), L.** The Science of Soap Films and Soap Bubbles. Cyril Isenberg. Tieto Ltd, 1978, xii + 188 pp, $15 (P). [ISBN: 0-905028-02-3] A fascinating exploration of the physical and mathematical principles of soap films--surface tension, interference phenomena, Steiner problems, minimal surfaces, Laplace-Young equation, calculus of variations--written for a general reader with an undergraduate background in science. Includes 12 pages of beautiful color plates, 6 appendices giving proofs of important results, and a well-organized bibliography of both popular and technical references. LAS

Applications (Social Science), P. Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems. A.G. Wilson. U of Calif Pr, 1981, 331 pp, $40. [ISBN: 0-520-04370-7] A survey of applications of catastrophe models to urban spatial structure (e.g., population flow, shopping center dynamics), preceded by "a lay guide" to the mathematics of catastrophe theory and followed by applications to other disciplines (biology, physical chemistry) selected as prototypes for new models of urban systems. Appendices treat more specialized topics. Prerequisite: thorough background in differential equations. LAS

Applications (Social Science), P, L.** Mathematical Methods in Social Science. David J. Bartholomew. Handbook of Appl. Math., Guidebook 1. Wiley, 1981, ix + 153 pp, $26.50; $14.50 (P). [ISBN: 0-471-27932-3; 0-471-27933-1] First Guidebook to a particular profession--sociology and its near neighbors--in Wiley's Handbook of Applicable Mathematics. Four chapters (on pattern and variation, surveys, multivariate methods, and dynamics of social systems) survey the use of mathematics in social science, with extensive cross references to the six-volume Handbook for exposition of the relevant mathematics. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Southern California Section

The fall meeting of the Southern California Section was held on Saturday, November 14, 1981, at the University of California at Santa Barbara. It was a joint meeting with AMS and the Southern California Section of SIAM.

Invited Lectures

"Some Remarkable New Sphere Packings," by N.J.A. Sloane, Bell Labs.
"Mathematical Methods of Economics," by Joel Franklin, Caltech.
"The Search for the Meaning of Independence," by Mark Kac, University of Southern California.

Short Presentations:

"A New Method of Primality Testing," by Leonard Adleman, University of Southern California.
"The Keller Conjecture and Some Related Problems," by Basil Gordon, University of California, Los Angeles.
"Egorychev's Proof of the van der Waerden Permanent Conjecture," by Henryk Minc, University of California, Santa Barbara.

Eastern Pennsylvania and Delaware Section

The Eastern Pennsylvania and Delaware Section met November 21, 1981 at Villanova University. There were 200 in attendance.

Invited Lectures:

"Does Mathematics Have Elements?" by Paul Halmos, Indiana University.
"Probability and the Approximation of Continuous Functions," by Jerry King, Lehigh University.
"The Proof of the Four Color Theorem," by Kenneth Appel, University of Illinois.

The Metropolitan New York Section

The fortieth Annual Meeting of the Metropolitan New York Section was held at Lehman College, City University of New York on Saturday, May 2, 1981, with approximately 120 persons in attendance.

Invited Lectures:

"Mathematical Models and Existence Theorems," by Dorothy L. Bernstein, Brown University.
"Geometric Versions of Some Algebraic Identities," by Howard Levi, Lehman College.

Short Presentations:

"A New Type Magic Latin 3-Cube of Order Ten," by Joseph Arkin and K. Singh, University of New Brunswick.
"Two Person Matrix Games: An Accessible Approach to Non-Cooperative Equilibria," by Martin E. Flashman, Bard College.
"Some Interesting Mathematical Models for Computer Solution," by Fred Marotto, Fordham University at Lincoln Center.
"The Role of the Computer in the Schools," a panel discussion by John Mineka, Lehman College; Ted Nelson (author and editor); and Randall Rustin (consultant).
"The Müntz-Szasz Theorem for Countable Sets," by James V. Peters, C.W. Post Center of Long Island University.
"Generalizations of Perfect Numbers with Group Applications," by Jay Schiffman, Kean College of New Jersey.
"Inequalities and the Evaluation of Integrals," by David Shelupsky, City College of New York.

Gregg Patruno of Stuyvesant High School received the Charles Salkind Award for the highest regional score in the M.A.A. High School Mathematics Contest. Mr. Patruno also received a trophy and a copy of International Mathematical Olympiads, 1959-1977 by Sam Greitzer.

Mark A. Weiss of Cooper Union received the Section Award for the highest regional score in the William Putnam Mathematical Competition. Mr. Weiss also received a copy of Topology of 3-Dimensional Fibered Spaces by H. Seibert.

4. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.
5. H. Hopf, Über Flächen mit einer Relation zwischen den Hauptkrümmungen, Math. Nachr., 4 (1951) 232–249.
6. R. Millman and G. Parker, Elements of Differential Geometry, Prentice-Hall, Englewood Cliffs, N.J., 1977.
7. R. Reilly, Applications of the Hessian operator in a Riemannian manifold, Indiana Univ. Math. J., 26 (1977) 459–472.
8. H. Weinberger, Remark on the preceding paper of Serrin, Arch. Rational Mech. Anal., 43 (1971) 319–320.

ELEMENTARY QUADRATURES OF ORDINARY DIFFERENTIAL EQUATIONS

LI HONG-XIANG (李鸿祥)
Shanghai Railway Institute, Shanghai 201803, China

Introduction. New types of ordinary differential equations which can be integrated in finite form are of wide interest and appeal to teachers of college mathematics and engineers who apply ordinary differential equations to their work. This MONTHLY has already published many notes on this subject (cf. [**5**]–[**9**], [**11**]). Here we give some new integrable types of first-order and second-order equations; many well-known classical (cf. [**1**]–[**4**]) and modern (cf. [**5**]–[**11**]) integrable types and their integrals are special cases of the results in this paper. We have introduced the concepts of the resolvent function, the characteristic constant and equation, and the discriminant. By using these concepts, the general solutions of many equations can be obtained by algebraic methods or formulas.

1. First-Order Equations.

THEOREM 1.1. *Suppose that $P, Q, F \in C$, $v \in C^1$, $v(x) \neq 0$ in some interval I. If the equality*

$$v' - P(x)v = kQ(x)v^2 \tag{1.1}$$

holds in I, where k is a constant, then the equation

$$y' + P(x)y = Q(x)F(yv(x)) \tag{1.2}$$

is integrable.

Proof. The transformation

$$y = \frac{u}{v} \tag{1.3}$$

converts (1.2) into the form

$$u' = v(x)Q(x)[F(u) + ku]$$

in which the variables are separable.

According to (1.1), after letting $k = 0$ and choosing $v = e^{\int P(x)\,dx}$, equation (1.2) becomes

$$y' + P(x)y = Q(x)F\left(y\exp\left(\int P(x)\,dx\right)\right), \tag{1.4}$$

and (1.3) becomes

$$y = ue^{-\int P(x)\,dx}. \tag{1.5}$$

It is obvious that, when $P(x) \equiv 0$, the equation (1.4) becomes

$$y' = Q(x)F(y)$$

in which the variables are separable, and (1.5) becomes $y = ue^{-A}$ (A = const.). That is, the given equation is already integrable; hence it is not necessary to make a transformation. When $P(x) = -1/x$, $Q(x) \equiv 1$, (1.4) becomes the so-called homogeneous equation

$$y' = \phi\left(\frac{y}{x}\right),$$

where $\phi(y/x) = y/x + F(y/x)$, and (1.5) becomes $y = xu$. When $F \equiv 1$, (1.4) becomes the linear equation

$$y' + P(x)y = Q(x).$$

When $F(u) = u^n$, $u = y\exp(\int P(x)\,dx)$. If we let $Q(x)\exp(n\int P(x)\,dx) = R(x)$, (1.4) becomes the Bernoulli equation

$$y' + P(x)y = R(x)y^n.$$

Thus some well-known integrable types of first-order ordinary differential equations are all particular cases of (1.4), and they can be reduced to separable equations by the unified transformation (1.5).

We also have the following corollaries.

COROLLARY 1.1. *Suppose that $P \in C$, Q, $R \in C^1$, and $R(x) \neq 0$. If $(1/P)(\ln(Q/R))' = 2$, then the Riccati equation*

$$y' + P(x)y = Q(x)y^2 + R(x) \tag{1.6}$$

can be reduced to one with separable variables by the transformation (1.5).

Proof. Substituting the given condition into (1.6), we get

$$y' + P(x)y = R(x)\left[ay^2\exp\left(2\int P(x)\,dx\right) + 1\right]$$

where a is a constant. This equation is a special case of (1.4) when $Q(x) = R(x)$ and $F(u) = au^2 + 1$.

This corollary is the result in [**5**].

COROLLARY 1.2. *If there exist constants c, k, and a function $v(x)$ such that $R(x) = cQ(x)v^2$, $v' + P(x)v = kR(x)$, then the Riccati equation* (1.6) *can be reduced to one with separable variables by* (1.3).

Proof. Substituting the given conditions into (1.6), we then obtain

$$y' - \frac{v'}{v}y = R(x)\left(\frac{y^2}{cv^2} - k\frac{y}{v} + 1\right)$$

which is a special case of (1.4) when $P(x) = -v'/v$, $Q(x) = R(x)$ and $F(u) = (1/c)u^2 - ku + 1$. Theorem 1.1. now applies.

This corollary is Theorem 1 in [**8**]. (Note that the functions f, g, h in [**8**] are respectively replaced by R, $-P$, Q here.) The result in [**5**] is a special case.

Generally, we have the following general corollary.

COROLLARY 1.3. *The first-order differential equation*

$$y' + P(x)y = Q(x)\left[\sum_{k=0}^{n} a_{n-k}y^k\exp\left(k\int P(x)\,dx\right)\right]^r, \tag{1.7}$$

where n is a natural number, and r and a_{n-k}, $k = 0, 1, \ldots, n$, are real constants, can be reduced to one with separable variables by (1.5).

EXAMPLE 1.1. The equation $x^{m(n-1)+n}y' = ay^n + bx^{(m+1)n}$ (see [3, p. 327, (1.189)]) can be rewritten in the form

$$y' - \frac{m+1}{x}y = x^m\left[a\left(\frac{y}{x^{m+1}}\right)^n - (m+1)\frac{y}{x^{m+1}} + b\right]$$

which is of the same type as (1.7), here $r = 1$, $P(x) = -(m+1)/x$, $Q(x) = x^m$, $a_0 = a$, $a_1 = a_2 = \cdots = a_{n-2} = 0$, $a_{n-1} = -(m+1)$, and $a_n = b$. Hence, according to Corollary 1.3, the given equation can be reduced to one with separable variables by (1.5), i.e., $y = x^{m+1}u$.

Obviously, if $r = 1$, $n = 2$, $a_0 = a$, $a_1 = -b$ and $a_2 = c$, the equation (1.7) becomes an integrable equation of Riccati type,

$$y' + P(x)y = Q(x)\left[ay^2\exp\left(2\int P(x)\,dx\right) - by\exp\left(\int P(x)\,dx\right) + c\right]. \tag{1.8}$$

Letting $r = 1, n = 3$ in (1.7), we obtain a class of integrable equations of Abel type of the first kind

$$y' + P(x)y = Q(x)\sum_{k=0}^{3} a_{n-k}y^k\exp\left(k\int P(x)\,dx\right). \tag{1.9}$$

EXAMPLE 1.2. The equation $x^{2n+1}y' = ay^3 + bx^{3n}$ (see [3, p. 326 (1.188)]) can be rewritten in the form

$$y' - \frac{n}{x}y = x^{n-1}\left[a\left(\frac{y}{x^n}\right)^3 - n\frac{y}{x^n} + b\right]$$

which is a special case of (1.9) when $P(x) = -n/x$, $Q(x) = x^{n-1}$, $a_0 = a$, $a_1 = 0$, $a_2 = -n$, and $a_3 = b$. Hence it is soluble by the transformation (1.5), i.e., $y = x^n u$.

It is obvious that, if α is a constant, then the equation

$$y' + P(x)y = y^\alpha Q(x)F\left(y\exp\left(\int P(x)\,dx\right)\right) \tag{1.10}$$

can also be reduced to one with separable variables by (1.5), and (1.4) is the special case when $\alpha = 0$.

THEOREM 1.2. *Suppose that $P, Q, F \in C$, $v, \phi \in C^1$, and $v(x) \neq 0$. If* (1.1) *holds, then the first-order differential equation*

$$y' + P(x)y = Q(x)F((y+\phi)v) - P(x)\phi - \phi' \tag{1.11}$$

can be reduced to one with separable variables by the transformation

$$y = \frac{u}{v} - \phi. \tag{1.12}$$

Proof. Substituting the given condition and (1.12) into the equation (1.11), we obtain

$$u' = vQ(x)[F(u) + ku]$$

whose variables are separable.

After letting $k = 0$ and choosing $v = \exp(\int P(x)\,dx)$ in (1.1), the equation (1.11) becomes the integrable equation

$$y' + P(x)y = Q(x)F\left((y+\phi)\exp\left(\int P(x)\,dx\right)\right) - P(x)\phi - \phi', \tag{1.13}$$

and (1.2) becomes

$$y = ue^{-\int P(x)\,dx} - \phi. \tag{1.14}$$

Obviously, the equation (1.4) is a special case of (1.13) when $\phi = 0$.

COROLLARY 1.4. *If there exist constants a, b and a function $V(x)$ such that $H = ahV^2$, $V' + gV = bH$, $H = f + (g/h)'$, then the Riccati equation*

$$y' = f(x) + g(x)y + h(x)y^2 \tag{1.15}$$

can be reduced to one with separable variables by $y = Vu - (g/h)$.

Proof. Letting $g(x) = P(x)$, $V = 1/v$, $H = Q(x)$ and $b = -k$, the condition $V' + gV = bH$ becomes (1.1). Since $H = ahV^2$ and $f = Q - (g/h)'$, (1.15) becomes

$$\begin{aligned} y' + Py &= \frac{Qv^2}{a}y^2 + 2Py + Q - \left(\frac{aP}{Qv^2}\right)' \\ &= Q\left[\frac{v^2}{a}\left(y + \frac{aP}{Qv^2}\right)^2 + 1\right] - \frac{aP^2}{Qv^2} - \left(\frac{aP}{Qv^2}\right)' \end{aligned}$$

which is the special case of (1.11) when $F(u) = (u^2/a) + 1$ and $\phi = aP/Qv^2$. According to Theorem 1.2, this equation can be reduced to one in which the variables are separable by (1.12) (notice that $V = 1/v$), i.e., $y = Vu - (g/h)$.

This corollary is Theorem 2 in [**8**]. The result in [**6**] is a special case.

We point out that, in [**3**], some 210 out of 367 equations, and in [**4**], about 460 out of 751 equations, which are of first degree in y', are all special cases of (1.4), (1.10) or (1.13) in the present paper.

2. Equations of Riccati Type.

THEOREM 2.1. *Let $Q \in C$, $G \in C^1$, $G(x) \neq 0$, and let a, b, c be real constants, and $a \neq 0$. Then the equation of Riccati type*

$$y' - \frac{G'(x)}{G(x)}y = Q(x)\left[ay^2 - bG(x)y + cG^2(x)\right] \tag{2.1}$$

is integrable, and its general solution is

$$\text{(i)}\quad y = G(x)\left[\omega\tan\left(a\omega\int G(x)Q(x)\,dx + A\right) + \frac{b}{2a}\right] \quad \text{when } b^2 < 4ac, \tag{2.2}$$

$$\text{(ii)}\quad y = -G(x)\left(\frac{1}{a\int G(x)Q(x)\,dx + A} - \frac{b}{2a}\right) \quad \text{when } b^2 = 4ac, \tag{2.3}$$

$$\text{(iii)}\quad y = -G(x)\left[\tau\tanh\left(a\tau\int G(x)Q(x)\,dx + A\right) - \frac{b}{2a}\right] \quad \text{when } b^2 > 4ac$$

$$\text{or}\quad y = -G(x)\left[\tau\coth\left(a\tau\int G(x)Q(x)\,dx + A\right) - \frac{b}{2a}\right], \tag{2.4}$$

where

$$\omega = \sqrt{\frac{c}{a} - \frac{b^2}{4a^2}}, \quad \tau = \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}, \tag{2.5}$$

and A is a constant of integration. (*The definitions of these constants will not be repeated later.*)

Proof. In fact, the equation (2.1) is a special case of (1.4) when $P(x) = -G'(x)/G(x)$, $Q(x)$ is replaced by $Q(x)G^2(x)$, $F(u) = au^2 - bu + c$. According to Theorem 1.1, the transformation (1.5), i.e.,

$$y = G(x)u \tag{2.6}$$

reduces (2.1) to the integrable equation

$$u' = G(x)Q(x)(au^2 - bu + c)$$

whose variables are separable. After integrating it and substituting the expression of u into (2.6),

we obtain (2.2), (2.3), and (2.4) when $b^2 < 4ac$, $b^2 = 4ac$, and $b^2 > 4ac$, respectively. This completes the proof.

COROLLARY 2.1. *If in* (1.15) $f = C_0^2 h \exp(2\int g\,dx)$, *where* C_0 *is a constant, and* f, g, h *are functions of* x, *then the general solution is*

(i) $y = \sqrt{\frac{f}{h}}\tan\left(\int \sqrt{fh}\,dx + A\right)$ *when* $fh > 0$,

(ii) $y = -\sqrt{-\frac{f}{h}}\tanh\left(\int \sqrt{-fh}\,dx + A\right)$ *when* $fh < 0$.

Proof. Substituting the given conditions into (1.15), and multiplying both sides of (1.15) by h, we obtain

$$(hy)' - \frac{1}{2}\left(\frac{f'}{f} + \frac{h'}{h}\right)(hy) = (hy)^2 + fh.$$

This is an equation of the type (2.1). If $fh > 0$, then $G(x) = \sqrt{fh}$, $Q(x) \equiv 1$, $a = 1$, $b = 0$, $c = 1$, and $b^2 < 4ac$. By using (2.2), we obtain (i). If $fh < 0$, then $G(x) = \sqrt{-fh}$, $Q(x) \equiv 1$, $a = 1$, $b = 0$, $c = -1$, and $b^2 > 4ac$. By using (2.4), we obtain (ii).

This is the result in [**3**, p. 23, §1, 4.9(c)]. Here f and h have been interchanged, and an error in sign in (ii) has been corrected.

We point out that the theorem in [**7**] is a special case of Theorem 2.1 in the present paper. In fact, under the condition

$$\frac{g + h'/(2h) - f'/(2f)}{\sqrt{fh}} = c'$$

where c' is a constant, (1.15) can be rewritten in the form

$$y' - \frac{1}{2}\left(\frac{f'}{f} - \frac{h'}{h}\right)y = h\left(y^2 + c'\sqrt{\frac{f}{h}}\,y + \frac{f}{h}\right)$$

which is of the same type as (2.1). Here $G(x) = \sqrt{f/h}$, $Q(x) = h$, $a = 1$, $b = -c'$, $c = 1$.

In addition. letting $G(x) = v$, $Q(x) = ch$, $a = 1/c'$, $b = k$, $c = 1$ in (2.1), we can obtain the result in Theorem 1 in [**8**] again.

REMARK. It is easy to see that the expressions (2.2) and (2.4) of the general solution of (2.1) can be unified as

$$y = G(x)\left[\omega\tan\left(a\omega\int G(x)Q(x)\,dx + A\right) + \frac{b}{2a}\right] \qquad \text{when } b^2 \neq 4ac, \tag{2.2'}$$

because, when $b^2 > 4ac$, we have

$$\omega = i\tau,$$

$$\omega\tan\left(a\omega\int G(x)Q(x)\,dx + A\right) = i\tau\tan\left(ia\tau\int G(x)Q(x)\,dx + A\right)$$

$$= -\tau\tanh\left(a\tau\int G(x)Q(x)\,dx + A'\right),$$

where $i = \sqrt{-1}$, $A' = -iA$.

THEOREM 2.2. *Let* $Q \in C$, $G, E \in C^1$, $G(x) \neq 0$, *and let* a, b, c *be real constants. Then the equation of Riccati type*

$$u' - \frac{G'}{G}u = Q\left[a(u+E)^2 - b(u+E)G + cG^2\right] + \frac{G'}{G}E - E' \tag{2.7}$$

is integrable, and its general solution is $u = y - E$ *where* y *is the general solution of* (2.1).

Proof. In fact, the transformation $u = y - E$ converts (2.7) into (2.1).

Note that the equation (2.7) is also a special case of (1.13) when $P(x) = -G'/G$, $Q(x)$ is replaced by $Q(x)G^2(x)$, $\phi = E$ and $F(w) = aw^2 - bw + c$. Hence, Theorem 2 in [**8**] is a corollary of Theorem 2.2 here.

From Theorem 2.2, we can derive several classical results obtained by Euler (1763, 1764), Weyr and Picard (1875–1877), namely, if a particular solution of the Riccati equation (1.15) is known, the general solution can be obtained by two quadratures; if two particular solutions are known, the general solution is obtainable by a single quadrature; if three particular solutions are known, the general solution can be effected without a quadrature (see [**1**, p. 93]).

For example, suppose that three particular solutions $y_1 = \phi$, $y_2 = \psi$, $y_3 = \chi$ of (1.15) are known, and that, without loss of generality, $h(x) \equiv 1$. According to the hypotheses, we obtain

$$g = \frac{\phi' - \psi'}{\phi - \psi} - (\phi + \psi), \quad f = -\phi\frac{\phi' - \psi'}{\phi - \psi} + \phi\psi + \phi',$$

$$\phi - \psi = \frac{\chi' - \phi'}{\chi - \phi} - \frac{\chi' - \psi'}{\chi - \psi}, \left(\frac{\chi - \phi}{\chi - \psi}\right)' = \frac{(\chi - \phi)(\phi - \psi)}{\chi - \psi}.$$

Hence the equation (1.15) can be rewritten in the form

$$y' - \left(\frac{\phi' - \psi'}{\phi - \psi} + \frac{\psi' - \chi'}{\psi - \chi} + \frac{\chi' - \phi'}{\chi - \phi}\right)y = (y - \phi)^2 - \phi\left(\frac{\phi' - \psi'}{\phi - \psi} + \frac{\psi' - \chi'}{\psi - \chi} + \frac{\chi' - \phi'}{\chi - \phi}\right) + \phi'$$

which is a special case of (2.7) when

$$G = \frac{(\chi - \phi)(\phi - \psi)}{\chi - \psi}, Q \equiv 1, E = -\phi, a = 1, b = c = 0.$$

Since $\Delta = 0$, by using theorem 2.2 we obtain the general solution

$$y = \phi - \frac{\dfrac{(\chi - \phi)(\phi - \psi)}{\chi - \psi}}{\dfrac{\chi - \phi}{\chi - \psi} + A} \qquad \text{or} \qquad \frac{y - \psi}{y - \phi} : \frac{\chi - \psi}{\chi - \phi} = A'$$

where $A' = -A$.

In addition, the result obtained by Kourensky in 1926 (see [**3**, p. 23, §1, 4.9(a)]) can also be derived with the help of the method mentioned above.

Obviously, the following equations are all special cases of (2.1) or (2.7):

$$y' = y^2 - \frac{a_1}{x}y + \frac{a_2}{x^2} \qquad (a_1, a_2 \text{ are constants}), \tag{2.8}$$

$$y' = y^2 - \left(bG + \frac{G'}{G}\right)y + cG^2 + \left(bG + \frac{G'}{G}\right)', \tag{2.9}$$

$$y' = a(y + E)^2 - b(y + E) + c - E'. \tag{2.10}$$

THEOREM 2.3. *Let $Q \in C$, $M, N, F, H, G \in C^1$, $MH - NF \neq 0$, $G(x) \neq 0$, and let a, b, c be real constants. Then the equation of Riccati type*

$$u' = R(x)u^2 + \Psi(x)u + \Phi(x), \tag{2.11}$$

where

$$R(x) = \frac{1}{MH - NF}\left[aQM^2 - M'N + MN' + MN\left(\frac{G'}{G} - bGQ\right) + cQN^2G^2\right], \tag{2.12}$$

$$\Psi(x) = \frac{1}{MH - NF}\left[MH' - M'H + FN' - F'N + (MH + NF)\left(\frac{G'}{G} - bGQ\right)\right.$$
$$\left. + 2aQMF + 2cQNHG^2\right], \tag{2.13}$$

$$\Phi(x) = \frac{1}{MH - NF}\left[aQF^2 - F'H + FH' + FH\left(\frac{G'}{G} - bGQ\right) + cQH^2G^2\right], \tag{2.14}$$

is integrable, and its general solution is $u = (-Hy + F)/(Ny - M)$ *where* y *is the general solution of* (2.1).

Proof. Substituting $u = (-Hy + F)/(Ny - M)$ into (2.11), we obtain the equation (2.1).

Obviously, (2.1) and (2.7) are all special cases of (2.11).

We call a, b, c the characteristic constants of (2.11), Q, G, etc., the resolvent functions, and $\Delta = b^2 - 4ac$ the discriminant. Therefore, to solve any Riccati equation which can be reduced to the type (2.1), (2.7), or (2.11), it is enough to determine the resolvent functions Q, G, etc., and the characteristic constants a, b, c. According to the sign of the discriminant Δ and corresponding Theorem, we can obtain the general solution of the given equation. This unifies and formalizes the solution of several classes of Riccati equations.

EXAMPLE 2.1. The equation $xy' + x^m + (n-m)/(2)y + x^n y^2 = 0$ ([**4**, p. 238, (184)]) can be rewritten in the form

$$y' - \frac{m-n}{2x}y = -x^{n-1}\left(y^2 + x^{m-n}\right)$$

which is of the same type as (2.1). Here the resolvent functions are $G(x) = x^{(m-n)/2}$, $Q(x) = -x^{n-1}$, and the characteristic constants are $a = 1, b = 0, c = 1$. Since $\Delta = -4 < 0$, $\omega = 1$, according to Theorem 2.1, we obtain the general solution

$$y = x^{(m-n)/2}\tan\left[-\frac{2}{m+n}x^{(m+n)/2} + A\right]$$

by using (2.2).

EXAMPLE 2.2. The equation $y' = 1 + x(2 - x^3) + (2x^2 - y)y$ ([**4**, p. 228, (48)]) can be rewritten in the form

$$y' = -\left(y - x^2\right)^2 + 1 - (-x^2)'$$

which is a special case of (2.7) (or (2.10)) when $G = Q \equiv 1$, $E = -x^2$, $a = -1$, $b = 0$, $c = 1$. Since $\Delta = 4 > 0$, $\tau = 1$, according to Theorem 2.2 we can use (2.4) to obtain the general solution

$$y = x^2 + \tanh(x + A).$$

We point out that all the integrable equations of Riccati type in [**3**] and [**4**] are of the same type as (2.1) or (2.7). Evidently many other integrable equations of Riccati type can be obtained by giving particular forms to the functions Q, G, M, N, H and F in (2.11).

3. Second-Order Linear Equations with Variable Coefficients.

THEOREM 3.1. *Let* $G \in C^1$, $G(x) \neq 0$, *and let* b, c *be real constants. Then the second-order homogeneous linear equation with variable coefficients*

$$v'' + \left(bG - \frac{G'}{G}\right)v' + cG^2 v = 0 \tag{3.1}$$

is integrable, and its general solution is

(i) $v = \left[C_1\cos(\omega\int G\,dx) + C_2\sin(\omega\int G\,dx)\right]\exp\left(-\frac{b}{2}\int G\,dx\right)$ *when* $b^2 < 4c$, (3.2)

(ii) $v = (C_1 + C_2\int G\,dx)\exp\left(-\frac{b}{2}\int G\,dx\right)$ *when* $b^2 = 4c$, (3.3)

(iii) $v = C_1\exp\left(\left(-\frac{b}{2} + \tau\right)\int G\,dx\right) + C_2\exp\left(\left(-\frac{b}{2} - \tau\right)\int G\,dx\right)$ *when* $b^2 > 4c$, (3.4)

where ω, τ *are of the same form as* (2.5) *when* $a = 1$; C_1, C_2 *are arbitrary constants.*

Proof. We make the substitution

$$v' = -yv. \tag{3.5}$$

This reduces (3.1) to (2.1) when $Q(x) \equiv 1$ and $a = 1$. If we substitute the solutions (2.2), (2.3), (2.4) of (2.1) into (3.5), after integrating and simplifying, we obtain (3.2), (3.3), (3.4), respectively.

REMARK. The expressions (3.2) and (3.4) of the general solution of (3.1) can be unified in the form

$$v = \left[C_1 \cos\left(\omega \int G\,dx\right) + C_2 \sin\left(\omega \int G\,dx\right)\right]\exp\left(-\frac{b}{2}\int G\,dx\right) \quad \text{when } b^2 \neq 4c. \tag{3.2'}$$

Obviously, the second-order Euler homogeneous equation ([**2**, p. 86])

$$v'' + \frac{a_1}{x}v' + \frac{a_2}{x^2}v = 0 \tag{3.6}$$

and the Legendre linear equation ([**4**, p. 87])

$$v'' + \frac{a_1}{a + dx}v' + \frac{a_2}{(a + dx)^2}v = 0, \tag{3.7}$$

where a_1, a_2, a, d are constants, are special cases of (3.1) when $G = 1/x$, $b = a_1 - 1$, $c = a_2$ and $G = 1/(a + dx)$, $b = a_1 - d$, $c = a_2$, respectively. In addition, the linear equation with constant coefficients

$$v'' + bv' + cv = 0 \tag{3.8}$$

is a special case of (3.1) when $G \equiv 1$.

REMARK. (3.6) can be reduced to an integrable Riccati equation (2.8) by (3.5).

THEOREM 3.2. *Let $G, F \in C^1$, $G(x) \neq 0$, and let b, c be real constants. Then the second-order linear differential equation*

$$u'' + \left(bG - 2F - \frac{G'}{G}\right)u' + \left[F^2 - F^1 - F\left(bG - \frac{G'}{G}\right) + cG^2\right]u = 0 \tag{3.9}$$

is integrable, and its general solution is $u = v\exp(\int F\,dx)$ where v is the general solution of (3.1).

Proof. The transformation

$$u' = -(y - F)u \tag{3.10}$$

reduces (3.9) to (2.1) ($Q(x) \equiv 1$, $a = 1$). Substituting the general solution of (2.1) into (3.10), integrating, and simplifying the result, we complete the proof.

Obviously, (3.9) corresponds to (2.7) ($Q \equiv 1$, $a = 1$).

When $F = -G'/G$, the equation (3.9) becomes

$$u'' + \left(bG + \frac{G'}{G}\right)u' + \left[\left(bG + \frac{G'}{G}\right)' + cG^2\right]u = 0 \tag{3.11}$$

which corresponds to the integrable equation (2.9), and whose general solution is obviously $u = v/G$ where v is the general solution of (3.1).

When $G(x) \equiv 1$, (3.9) becomes

$$u'' + (b - 2F)u' + (F^2 - F' - bF + c)u = 0 \tag{3.12}$$

which corresponds to the integrable Riccati equation (2.10), whose general solution is $u = v\exp(\int F\,dx)$ where v is the general solution of (3.8).

From Theorem 3.2, we can also derive many classical results such as the conclusions in [**2**, p. 34], [**4**, pp. 88–89, B1, 4-1-1(a), 4-1-2(iii)], and [**3**, pp. 119–120, §6.25.1(b), (e)].

For example, if s is any particular solution of the equation $\{s, x\} = 2I$ where

$$\{s, x\} = \frac{s'''}{s'} - \frac{3}{2}\left(\frac{s''}{s'}\right)^2$$

is called the Schwartzian derivative, and the quantity

$$I = \frac{h}{f} - \frac{1}{4}\left(\frac{g}{f}\right)^2 - \frac{1}{2}\left(\frac{g}{f}\right)'$$

is the invariant of $fy'' + gy' + hy = 0$, then the function

$$u = \frac{1}{\sqrt{s'}}(C_1 + C_2 s) \tag{3.13}$$

is the general solution of the equation

$$u'' + Iu = 0. \tag{3.14}$$

In fact, substituting $I = \frac{1}{2}\{s, x\}$ into (3.14), we obtain

$$u'' + \left[\frac{s'''}{2s'} - \frac{3}{4}\left(\frac{s''}{s'}\right)^2\right]u = 0$$

which is a special case of (3.9) when $G(x) = s'$, $F = -s''/(2s')$, $b = c = 0$. Since $\Delta = 0$, the general solution (3.13) follows from Theorem 3.2 and (3.3).

From Theorem 3.2, we can also derive some results obtained by Zeitlin [9] in 1977. For instance, if $P \in C^2$, $P(x)P'(x) \neq 0$, $D \equiv (d/dx)$, then

$$u = \exp(P(x)) = \sum_{k=0}^{\infty} \frac{P^k}{k!} \qquad \text{and } U_n = \sum_{k=0}^{n} \frac{P^k}{k!}$$

are both solutions of the equation

$$\left(D - \frac{P''}{P'} - n\frac{P'}{P}\right)(D - P')u = 0 \tag{3.15}$$

([9, Theorem 2]).

Obviously, (3.15) can be rewritten in the form

$$u'' - \left(\frac{P''}{P'} + n\frac{P'}{P} + P'\right)u' + n\frac{P'^2}{P}u = 0$$

which is a special case of (3.9) when $G(x) = P'P^n e^{-P}$, $F(x) = P'$, $b = c = 0$, and $\Delta = 0$. According to Theorem 3.2 and by using (3.3), we obtain the general solution

$$u = e^P\left(C_1 + C_2'\int P'P^n e^{-P}\,dx\right) = C_1 e^P + C_2 \sum_{k=0}^{n} \frac{P^k}{k!}.$$

The equation in [11] is a special case of (3.15) when $P(x) = x$.

We point out that some 280 among 445 equations in [3] and about 285 among 596 equations in [4] are all of the same type as (3.1) or (3.9). In addition, the most general integrable equation given by Zbornik in [10]

$$y'' - \frac{f''}{f'}y' - \frac{f'^2}{F(f)}\frac{d^2F}{df^2}y = 0 \tag{3.16}$$

is also a special case of (3.9) when $G = f'/F^2(f)$, and F is replaced by $(d/dx)(\ln F(f))$, $b = c = 0$. Hence 57 out of 66 equations mentioned in [10] are all of the same type as (3.1) or (3.9).

THEOREM 3.3. *Let $Q, F, H \in C^1$, $M, N, G \in C^2$, $MH - NF \neq 0$, $G(x) \neq 0$, $R(x) \neq 0$, and let a, b, c be real constants. Then the second-order linear differential equation*

$$y'' - \left(\Psi + \frac{R'}{R}\right)y' + \Phi R y = 0, \tag{3.17}$$

where R, Ψ, Φ *are the same as* (2.12), (2.13), (2.14) *respectively, is integrable, and its general solution is*

$$y = B\exp\left(-\int Ru\,dx\right)$$

where u *is the general solution of* (2.11), *and* B *is an arbitrary constant.*

Proof. Making the transformation

$$y' = -Ruy \tag{3.18}$$

in (3.17), we obtain the integrable Riccati equation (2.11). Substituting its general solution into (3.18) and then integrating completes the proof.

Obviously, (3.1) and (3.9) are both special cases of (3.17); and the inhomogeneous linear equations obtained by adding a continuous function $f(x)$ to their right-hand sides are also integrable.

As in Section 2, we call a, b, c the characteristic constants of (3.17), Q, G, etc. the resolvent functions, $\Delta = b^2 - 4ac$ the discriminant, $ar^2 + br + c = 0$ the characteristic equation. This also reduces the solution of (3.1), (3.9), (3.11), (3.12) and (3.17) to formulas.

EXAMPLE 3.1. The equation $(1-x^2)^2y'' - 2x(1-x^2)y' - a^2y = 0$ where a is a positive number ([**4**, p. 373 (553)]) is a special case of (3.1). Here the resolvent function is $G(x) = 1/(1-x^2)$, the characteristic constants are $b = 0$, $c = -a^2$, the discriminant is positive: $\Delta = 4a^2 > 0$, $\tau = a$, and the characteristic equation $r^2 - a^2 = 0$ has two different roots $r_{1,2} = \pm a$. According to Theorem 3.1, and by using (3.4), we get the general solution

$$y = C_1\left(\frac{1+x}{1-x}\right)^{a/2} + C_2\left(\frac{1-x}{1+x}\right)^{a/2}.$$

EXAMPLE 3.2. $x^2y'' + xy' + (x^2 - \frac{1}{4})y = 0$ is a Bessel equation when $n = \pm\frac{1}{2}$. It is easy to verify that it is of the same type as (3.12). Here $F(x) = -1/(2x)$, $b = 0$, $c = 1$, $\Delta < 0$, $\omega = 1$. Thus its general solution is

$$y = \frac{1}{\sqrt{x}}(C_1\cos x + C_2\sin x).$$

4. Second-Order Nonlinear Equations. By using the results in Section 2 and Section 3, it is easy to prove the following theorems.

THEOREM 4.1. *Let the functions* Q, G *and the constants* a, b, c *satisfy the same hypotheses of Theorem* 2.1. *Then the second-order nonlinear equation*

$$v'' + \left(bGQ - \frac{G'}{G}\right)v' - aQv'^2 = cQG^2 \tag{4.1}$$

is integrable, and its general solution is $v = \int y\,dx + B$ *where* y *is the general solution of* (2.1).

THEOREM 4.2. *Let the functions* Q, G, E *and the constants* a, b, c *satisfy the hypotheses of Theorem* 2.2 *Then the second-order nonlinear equation*

$$v'' + \left(bGQ - 2aQE - \frac{G'}{G}\right)v' - aQv'^2 = Q(aE^2 + cG^2) - E' - E\left(bGQ - \frac{G'}{G}\right) \tag{4.2}$$

is integrable, and its general solution is $v = \int u\,dx + B$ *where* u *is the general solution of* (2.7).

THEOREM 4.3. *Let the functions* R, Ψ, Φ *satisfy the hypotheses of* (2.12), (2.13), (2.14), *respectively. Then the nonlinear equation*

$$v'' - \Psi(x)v' - R(x)v'^2 = \Phi(x) \tag{4.3}$$

is integrable, and its general solution is $v = \int u\,dx + B$, *where* u *is the general solution of* (2.11).

THEOREM 4.4. *Let* $L[v] \equiv v'' + r(x)v' + s(x)v$ *be the left side of* (3.1), (3.9) *or* (3.17). *Then the second-order nonlinear equation*

$$L[v] = \frac{k}{v^3}\exp\left(-2\int r(x)\,dx\right), \tag{4.4}$$

where k *is a constant, is integrable, and its general solution is*

$$v = \left(C_1 v_1^2 + C_2 v_2^2 + 2Av_1 v_2\right)^{1/2} \tag{4.5}$$

where v_1, v_2 *are a basis for the solution of* $L[v] = 0$, *and* A *is a constant determined by the expression*

$$C_1 C_2 - A^2 = \frac{k}{(v_1 v_2' - v_1' v_2)^2}\exp\left(-2\int r(x)\,dx\right).$$

(*Note that* $(v_1 v_2' - v_1' v_2)\exp(\int r(x)\,dx) = const.$)

EXAMPLE 4.1. The second-order nonlinear differential equation $xy'' + \alpha(xy' - y)^2 = \beta$, where α, β are real constants ([**3**, p. 564, (6.80)]), can be changed to

$$v'' + \frac{2}{x}v' + \alpha x^2 v'^2 = \frac{\beta}{x^2}$$

by the transformation $y \doteq xv$. This equation is a special case of (4.1). Here $G(x) = 1/x^2$, $Q(x) = x^2$, $a = -\alpha$, $b = 0$, $c = \beta$. Hence, letting $v' = z$, it can be reduced to an integrable Riccati equation which belongs to the type (2.1). Solving it and returning into original unknown function y, we can get the general solution of the given equation.

References

1. G. N. Watson, A Treatise on the Theory of Bessel Functions, 2nd ed., Cambridge University Press, London, 1952, Chap. 4.

2. G. Birkhoff and G.-C. Rota, Ordinary Differential Equations, 2nd ed., Wiley, New York, 1969.

3. E. Kamke, Differentialgleichungen: Lösungsmethoden und Lösungen, I. Gewöhnliche Differentialgleichungen, 9. Aufl., B.G. Teubner Stuttgart, 1977.

4. G. M. Murphy, Ordinary Differential Equations and Their Solutions, D. Van Nostrand, New York, 1960.

5. A. K. Rajagopal, On the generalized Riccati equation, this MONTHLY, 68 (1961) 777–779.

6. P. R. P. Rao, The Riccati differential equations, this MONTHLY, 69 (1962) 995.

7. J. L. Allen and F. M. Stein, On solutions of certain Riccati differential equations, this MONTHLY, 71 (1964) 1113–1115.

8. P. R. P. Rao and V. H. Ukidave, Some separable forms of the Riccati equation, this MONTHLY, 75 (1968) 38–39.

9. D. Zeitlin, On a class of ordinary linear differential equations having $\Sigma_{k=0}^{\infty} c_k x^k$ and $\Sigma_{k=0}^{n} c_k x^k$ as solutions, this MONTHLY, 84 (1977) 716–720.

10. J. Zbornik, Lösungsmethoden der homogenen linearen Differentialgleichungen zweiter Ordnung, Z. Angew. Math. Phys., 7 (1956) 64–74.

11. J. P. Hoyt, Remarks on a second order differential equation, this MONTHLY, 74 (1967) 436.

THE TENTH U.S.A. MATHEMATICAL OLYMPIAD

SAMUEL L. GREITZER
Department of Mathematics, Rutgers University, New Brunswick, NJ 08903

Participants in this examination, held May 5, 1981, were 150 top-ranking students in the 1981 Annual High School Mathematics Examination (AHSME), the competition from which Olympiad participants have always been selected. On March 10, 1981, 422,000 students took part in the AHSME.

The eight highest scorers in the Olympiad constituted the winners. In order of rank they were:

*Noam D. Elkies	Stuyvesant High School	New York, NY
*Gregg N. Patruno	Stuyvesant High School	New York, NY
Richard A. Stong	Albermarle High School	Charlottesville, VA
Benji N. Fisher	Bronx High School of Science	Bronx, NY
Jeremy D. Primer	Columbia High School	Maplewood, NJ
Brian R. Hunt	Montgomery Blair High School	Silver Spring, MD
David S. Yuen	Lane Technical High School	Chicago, IL
James R. Roche	Hill-Murray High School	St. Paul, MN

(*Indicates tied score.)

The questions of the Tenth Olympiad follow:

1. The measure of a given angle is $180°/n$, where n is a positive integer not divisible by 3. Prove that the angle can be trisected by Euclidean means (straight edge and compasses).
2. Every pair of communities in a county are linked directly by exactly one mode of transportation: bus, train or airplane. All three modes of transportation are used in the county with no community being serviced by all three modes and no three communities being linked pairwise by the same mode. Determine the maximum number of communities in this county.
3. If A, B, and C are the measures of the angles of a triangle, prove that
$$-2 \leqslant \sin 3A + \sin 3B + \sin 3C \leqslant 3\sqrt{3}/2$$
and determine when equality holds.
4. The sum of the measures of all the face angles of a given convex polyhedral angle is equal to the sum of the measures of all its dihedral angles. Prove that the polyhedral angle is a trihedral angle. *Note*: A convex polyhedral angle may be formed by drawing rays from an exterior point to all points of a convex polygon.
5. If x is a positive real number and n is a positive integer, prove that
$$[nx] \geqslant \frac{[x]}{1} + \frac{[2x]}{2} + \frac{[3x]}{3} + \cdots + \frac{[nx]}{n},$$
where $[t]$ denotes the greatest integer less than or equal to t. For example, $[\pi] = 3$ and $[\sqrt{2}] = 1$.

Though the number of participants in the Olympiad this year was larger than the number usually participating, there was no increase in the number of high scores. The distribution this year, with a score of 150 the maximum, shows:

Range of Scores	*No. of Students*
0	24
1–20	75
21–40	30
41–60	14
61–80	7

For comparison, the ratio of the scores 60 or better to the number taking the test in previous years was:

1972	10/100	1975	6/103	1978	12/108
1973	4/107	1976	8/94	1979	7/98
1974	9/149	1977	8/107	1980	8/120

Adding more participants this year appeared only to increase the number of low scores.

The fact that two-thirds of the students scored 20 points or below shows that the HSME appears to determine the talented students, while the USAMO appears to locate gifted students. This is quite natural, since this is what the two tests are expected to do.

As for results on the types of problems, they are the same as those in previous Olympiads. Students showed a marked weakness in geometry, a weakness in trigonometry, and insufficient practice in such areas as elements of number theory. It is a pity that schools do not present these subjects to their talented and gifted students.

A new event this year: a girl requested that the Olympiad be given to her in Spanish, and this was done. Consideration should be given in the future to giving the examination in Spanish and French. Incidentally, there were eight girls who participated in the Olympiad this year.

The Award Ceremony has been made possible, ever since its inception, by grants from IBM. Various prizes were contributed by the Mathematical Association of America, by several publishing houses including Academic Press, Birkhauser Boston, W. H. Freeman, Springer-Verlag, Wadsworth, and John Wiley, and by the Hewlett-Packard Corporation.

Solutions of the problems of the Tenth USAMO will be available from Professor Walter E. Mientka, University of Nebraska, Lincoln, NE 68588.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

AN ANALOG OF BOLZANO'S THEOREM FOR FUNCTIONS OF A COMPLEX VARIABLE

MAU-HSIANG SHIH

Department of Mathematics, Chung-Yuan College of Science and Engineering, Chung-Li, Taiwan

The well-known Bolzano theorem [**1**, p. 85] states that if f is a real-valued continuous function on a closed interval $[a, b]$ and $f(a)f(b) < 0$, then f has a zero in (a, b). Without loss of generality we can suppose that $a < 0 < b$ and $f(a) < 0 < f(b)$. Then the condition $f(a)f(b) < 0$ becomes $xf(x) > 0$ for $x \epsilon \partial I$ where ∂I denotes the boundary of the interval $I = (a, b)$. The conclusion is that f has at least one zero in I.

It is a remarkable fact that this form of Bolzano's theorem extends to analytic functions of a complex variable and, indeed, with a stronger conclusion. We shall establish the following:

THEOREM 1. *Let Ω be a bounded domain of the z plane containing the origin. Let f be analytic in Ω and continuous in $\bar{\Omega}$, and suppose* $\operatorname{Re} \bar{z}f(z) > 0$ *for $z \epsilon \partial\Omega$. Then f has exactly one zero in Ω.*

For the proof let $g(z) = cz - f(z)$ where c is defined by

$$c = \frac{\inf \operatorname{Re} \bar{z} f(z)}{\sup |z|^2}, \qquad z \epsilon \partial\Omega.$$

Then on $\partial\Omega$ we have

$$|f(z)|^2 - |g(z)|^2 = |f(z)|^2 - |cz - f(z)|^2 = 2c \operatorname{Re} \bar{z} f(z) - c^2 |z|^2 > 0.$$

Thus $|f(z)| > |g(z)|$ on $\partial\Omega$; by Rouché's theorem $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside Ω, and the theorem follows.

We conclude with two supplementary remarks. At first glance it is not obvious that Rouché's theorem applies, since the usual proof [**2**] depends on Cauchy's formula and we have assumed no regularity of $\partial\Omega$. The desired extension can be obtained, however, by applying Rouché's theorem in its usual form to suitable sets of squares contained (with their closures) in Ω.

References

1. T. M. Apostol, Mathematical Analysis, 2nd ed., Addison-Wesley, Reading, Mass., 1974.

2. N. Levinson and R. M. Redheffer, Complex Variables, Holden-Day, San Francisco, Calif., 1970.

RINGS OF SETS ARE REALLY RINGS

J. B. WILKER

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada, M5S 1A1

A ring of subsets of X is a collection R of subsets A, B, etc., that is closed under unions $A \cup B$ and differences $A \setminus B$ and hence under intersections $A \cap B = A \setminus (A \setminus B)$ and symmetric differences $A \Delta B = (A \setminus B) \cup (B \setminus A)$. A number of texts on measure theory (see, for example, [**1**, p. 3] or [**3**, p. 22]) point out that a ring of sets becomes a ring in the algebraic sense if addition is defined by symmetric difference $A + B = A \Delta B$ and multiplication is defined by intersection $AB = A \cap B$.

A direct proof of this result is fairly tedious. To give a fast proof we recall that the characteristic function of a set A is defined by $\chi_A(x) = 0$ if $x \notin A$ and $\chi_A(x) = 1$ if $x \in A$. In measure theory the values 0 and 1 are usually taken to be real numbers, but we can also interpret them as elements of the 2-element field Z_2. With this unorthodox interpretation it is easy to see that $\chi_{A \Delta B} = \chi_A + \chi_B$ and $\chi_{A \cap B} = \chi_A \chi_B$. The result follows immediately because these equations show that R is isomorphic to a subring of Z_2^X where Z_2^X is the ring of Z_2-valued functions on X with pointwise addition and multiplication. We conclude by observing that if R is equal to the power set of X then it is isomorphic to the full ring Z_2^X.

A referee has kindly drawn my attention to two facts. On the one hand, the proof given above is already mentioned in [**2**, p. 137, Problem 18]. On the other hand, several recent books on ring theory state the result but continue to indicate the more laborious direct proof.

References

1. S. K. Berberian, Measure and Integration, Macmillan, New York, 1965.

2. N. Bourbaki, Algèbre, Chapitre 1, Structures Algébriques, Hermann, Paris, 1942.

3. P. R. Halmos, Measure Theory, Van Nostrand, Princeton, 1964.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by July 31, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2932. *Proposed by Henry E. Fettis, Mountain View, CA.*

Set

$$S_n(b) = \sum_{k=0}^{\left[\frac{n+b}{2}\right]} \frac{(n-2k+b)^{n-1}}{k!\,(n-k)!}.$$

Prove that $S_n(b) = S_n(-b)$.

E 2933. *Proposed by Doug Hensley, Texas A & M University.*

Show that for $1 \leqslant k \leqslant n$

$$2\binom{n}{k}^{1/n} \geqslant \binom{n}{k-1}^{1/n} + \binom{n}{k+1}^{1/n}.$$

E 2934. *Proposed by R. P. Boas, Northwestern University.*

Let f be a real-valued continuous function on $[0,1]$ and let h be a number between 0 and 1. Suppose the average of f over each subinterval of $(0,1)$ of length h is less than 1. Can the average of f over $[0,1]$ be greater than 1?

E 2935.* *Proposed by Clark Kimberling, University of Evansville.*

Let n be a prime greater than 10, and let $\{\quad\}$ denote fractional part. Prove or disprove that the number $S(m)$ of positive integer solutions j of the inequality $\{mj/n\} \geqslant 2j/n$ satisfies

$$S\left(\frac{n+1}{2}\right) \leqslant S(m) \leqslant S\left(\frac{n-1}{2}\right)$$

for $m=5,6,7,\ldots,n-5$.

E 2936. *Proposed by H. Kestelman, University College, London.*

An $n\times n$ complex matrix $A=[a_{ij}]$ is cross-diagonal if $a_{ij}=0$ whenever $i+j\neq n+1$. Find the condition that the eigenvectors of A span $\mathbb{C}^n$, the entire n-vector space.

E 2937. *Proposed by A. N. Philippou and G. D. Stamatelos, University of Patras, Greece.*

Suppose γ_n is a sequence such that $\gamma_{2n}\to\alpha$, $\gamma_{2n+1}\to\beta$. Show that

$$\frac{1}{n}\sum_{j=1}^{n}\frac{j}{j+1}\gamma_j\to\frac{1}{2}(\alpha+\beta).$$

SOLUTIONS OF ELEMENTARY PROBLEMS

A Product of Powers of an Integer

E 2852 [1980, 672]. *Proposed by Jan Mycielski, University of Colorado.*

For any positive integer n let $\omega(n)$ be the product of all positive integers k such that n is a power of k. Prove that

$$\prod_{n=1}^{\infty}\left(\frac{\omega(n)}{n+1}\right)^{1/n}=1.$$

I. *Solution by L. Van Hamme, Vrije Universiteit, Brussels, Belgium.* We prove the assertion in the form

$$\sum_{n=1}^{\infty}\frac{\log\omega(n)}{n}=\sum_{n=1}^{\infty}\frac{\log(n+1)}{n}.$$

Since $\omega(n)=\prod_{k^a=n}k$, $\log\omega(n)=\sum_{k^a=n}\log k$, we have

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{\log\omega(n)}{n}&=\sum_{n=1}^{\infty}\frac{1}{n}\sum_{k^a=n}\log k\\&=\sum_{k=1}^{\infty}\log k\sum_{k^a=n}\frac{1}{n}=\sum_{k=2}^{\infty}\log k\sum_{a=1}^{\infty}\frac{1}{k^a}\\&=\sum_{k=2}^{\infty}\frac{\log k}{k-1}=\sum_{n=1}^{\infty}\frac{\log(n+1)}{n}.\end{aligned}$$

We used absolute convergence in changing the order of summation.

II. *Solution by Sidney Heller, Brookhaven National Lab.* The integer $\kappa\geqslant 2$ occurs in $\omega(n)$ if and only if $n=\kappa^r$ (where $r=1,2,\ldots$). Hence the power of κ which occurs in the numerator of the product is

$$\kappa^\gamma=\kappa^{1/(\kappa-1)},\quad\text{where }\gamma=\sum_{r=1}^{\infty}\frac{1}{\kappa^r}.$$

But $\kappa^{1/(\kappa-1)}$ is the power of κ which occurs in the denominator, and since this is true for all $\kappa\geqslant 2$, the value of the product is 1.

Also solved by Ron Adin (student, Israel), R. Balasubramian & James Hafner & Maruti R. Murty, Robert Breusch, Nick Franceschine III, Carl Hurd, David Iny (student), O. P. Lossers (Netherlands), L. E. Mattics, Claudia Spiro, Alan H. Stein, Dean C. Wills (student), and the proposer.

SOLUTIONS OF ADVANCED PROBLEMS

Collinearity Preserving Maps in Affine Spaces

5790 [1971, 410]. *Proposed by D. E. Daykin, The University, Reading, England.*

Find all nontrivial maps $f: R^2 \to R^2$ such that whenever a, b, c are collinear, then $f(a), f(b), f(c)$ are collinear.

Solution by David S. Carter, Oregon State University, and Andrew Vogt, Georgetown University. Study of this deceptively innocent-looking problem grew into a full research project, and led us to consider the more general problem of characterizing all collinearity-preserving functions from one affine or projective Desarguesian plane into another. Our eventual solution to the latter problem is summarized in [**1**] and presented in detail in [**2**]. Various portions of the solution have been obtained independently by Klingenberg, Skornyakov, Radó, Garner, Barnes, Gibbons, and (jointly) Hales and Straus. For references, see [**2**].

The collinearity-preserving functions in the affine plane R^2 are: (i) trivial functions of collapsed, axial, or radial type; (ii) certain nontrivial functions with range containing a triangle but no quadrangle; and (iii) the usual affine transformations.

The functions in (i) are easily described: a collapsed function maps the whole plane to points on a single line, an axial one maps all points off one line to a single point, and a radial one maps all lines off one point to a single line. The functions in (ii), on the other hand, have rather intricate definitions in terms of proper valuation subrings of R and fractional ideals of R with respect to these rings. When one of these functions has a three-point range, it induces a partition of R^2 into three sets such that every line meets exactly two of the sets and none of the three sets contains a continuum. The valuation rings required for the definition of these functions exist according to Zorn's Lemma, but they are not measurable. Hence, if the Axiom of Choice is not used, the prospect of exhibiting these functions explicitly is at least as slim as the prospect of exhibiting nonmeasurable subsets of R.

A key step in our solution of the general problem is an extension theorem which asserts that, with the exception of certain embeddings defined on planes of order 2 and 3, every collinearity-preserving function from one affine Desarguesian plane into another can be extended to a collinearity-preserving function between enveloping projective planes. The extension theorem permits us to pass to projective planes and apply the fundamental theorem of projective geometry. The latter theorem characterizes collinearity-preserving bijections in terms of semilinear transformations and associated division ring isomorphisms. If valuation-theoretic places are substituted for division ring isomorphisms, the fundamental theorem can be modified to include collinearity-preserving functions with range containing a quadrangle, the resulting functions in many cases being neither onto nor one-to-one. However, since a place of R into R must be the identity map, it follows that the only collinearity-preserving functions from R^2 into R^2 with range containing a quadrangle are the affine transformations.

1. D. S. Carter and A. Vogt, Collinearity-preserving functions between Desarguesian planes, Proc. Nat. Acad. Sci. U.S.A., 77 (1980) 3756–3757.

2. ———, Collinearity-preserving functions between Desarguesian planes, Memoirs Amer. Math. Soc., no. 235 (1980).

Projectives in *-Rings without Upper Bound

6286 [1980, 65]. *Proposed by F. Gerrish, The Polytechnic, Kingston-upon-Thames, England.*

Let R be a *-ring, i.e., a ring (not necessarily with unity element) together with a conjugation $*: R \to R$ such that $(x+y)^* = x^* + y^*$, $(xy)^* = y^*x^*$, and $x^{**} = x$ (for all x, y in R). An element x in R is called *projective* if x is idempotent and self-conjugate, i.e., $xx = x = x^*$. Let P be the set of all projective elements in R. Then the relation $\leqslant$ on P defined by "$x \leqslant y$ if and only

if $x = xy$" is a partial order on P. Give an example, preferably simple, of a *-ring R for which $(P, \leqslant)$ is *not* a lattice.

(This problem arose from a remark, near the foot of page 49, in the article, "The Current Interest in Orthomodular Lattices," by S. S. Holland, in *Trends in Lattice Theory*, by J. C. Abbott, Van Nostrand, 1970.)

Solution by A. A. Jagers, Technische Hogeschool Twente, Enschede, Netherlands. Let S be the right zero semigroup on the set $\{e, f\}$, $e \neq f$ (product rule: $xy = y$). Let $\mathbb{Z}_2[S]$ be the semigroup algebra of S over $\mathbb{Z}_2$, and let R be the direct sum of $\mathbb{Z}_2[S]$ and its opposite ring, equipped with the exchange involution $(x, y)^* = (y, x)$. Then R is a *-ring without unity, and $P = \{(0,0), (e, e), (f, f)\}$, but (e, e) and (f, f) have no common upper bound.

Solution by Miroslav D. Ašić, University of Belgrade, Yugoslavia. Jagers' solution can be described as a 16-element *-algebra over $\mathbb{Z}_2$ with basis $\{p, q, r, s\}$, involution determined by $p^* = r$, $q^* = s$, and multiplication as given in Table 1.

	p	q	r	s
p	p	q	0	0
q	p	q	0	0
r	0	0	r	r
s	0	0	s	s

TABLE 1

	p	q	r	s
p	p	q	p	q
q	p	q	p	q
r	r	s	r	s
s	r	s	r	s

TABLE 2

Ašić constructs another nonisomorphic 16-element *-algebra over $\mathbb{Z}_2$ with basis $\{p, q, r, s\}$, involution determined by $p^* = p$, $s^* = s$, $q^* = r$, and multiplication as given in Table 2. Here $P = \{0, p, s\}$, and p and s have no common upper bound.

Also solved by F. S. Cater, with an infinite dimensional matrix ring. In a recent preprint, Y. Kato (Japan) states that the projections in the Calkin algebra do not form a lattice.

Inequality between Products

6294 [1980, 309]. *Proposed by Murray S. Klamkin, University of Alberta.*

If $n = n_1 + n_2 + \cdots + n_r$ where $n_r \geqslant 0$, prove that

$$\frac{n^n}{\Pi n_i^{n_i}} \geqslant \frac{\Gamma(1+n)}{\Pi\Gamma(1+n_i)} \geqslant \frac{(n+1)^{n+1}}{\Pi(n_i+1)^{n_i+1}}.$$

Solution by the proposer. The two inequalities will follow easily if we can show that the two functions

$$F(x) = \log(x^x/\Gamma(1+x)) \qquad \text{and } G(x) = \log\left(\Gamma(1+x)/(1+x)^{(1+x)}\right)$$

are superadditive, e.g., $F(x+y) \geqslant F(x) + F(y)$ for $x \geqslant 0$. Since $F(0) = G(0) = 0$, it suffices to show that both $F(x)$ and $G(x)$ are convex or equivalently that $F''(x) \geqslant 0$, $G''(x) \geqslant 0$ (see D. S. Mitrinović, *Analytic Inequalities*, Springer-Verlag, Heidelberg, 1970, pp. 22–23).

$$F'(x) = 1 + \log x - \Gamma'(1+x)/\Gamma(1+x) = 1 + \log x - \psi(1+x),$$

$$F''(x) = 1/x - \psi'(1+x) = \frac{1}{x} - \left\{\frac{1}{(1+x)^2} + \frac{1}{(2+x)^2} + \cdots\right\}.$$

By the integral test,

$$\psi'(1+x) \leqslant \frac{1}{(1+x)^2} + \int_{1+x}^{\infty} \frac{dt}{t^2} = \frac{1}{(1+x)^2} + \frac{1}{(1+x)} = \frac{2+x}{(1+x)^2} < \frac{1}{x}.$$

Thus, $F(x)$ is superadditive and

$$\sum F(n_k) \leqslant F\left(\sum n_i\right) = F(n),$$

which establishes the left-hand half of (1).

$$G'(x) = \psi(1+x) - 1 - \log(1+x),$$
$$G''(x) = \psi'(1+x) - 1/(1+x).$$

To show $G''(x) \geqslant 0$, first note that

$$G''(x) = \int_0^\infty \frac{te^{-(x+1)t}\,dt}{1-e^{-t}} - \int_0^\infty e^{-(x+1)t}\,dt = \int_0^\infty \frac{(t-1+e^{-t})}{1-e^{-t}} e^{-(x+1)t}\,dt;$$

so $G''(x) \geqslant 0$, since $e^{-t} \geqslant 1-t$ for $t \geqslant 0$. Thus, $G(x)$ is also superadditive which gives the right-hand half of (2).

It is to be noted that the case of the left-hand inequality, when the n_i's are nonnegative integers, reduces to a problem of Leo Moser (Math. Mag., 31 (1957) 113). The neat solution by Chih-yi Wong was to first write $(n_1 + n_2 + \cdots + n_r)^n = n^n$ and then to note that each term of the multinomial expansion of the left-hand expression is less than n^n.

Also solved by Otto G. Ruehr and incompletely by both N. Franceschine and L. Kuipers since they assumed that the n_i were integers.

Editorial Note: The solution by Ruehr suggests a further problem: Determine

$$\sup\left\{\alpha : \sum f_\alpha(t_i) \geqslant f_\alpha(t), \quad \text{whenever } t = t_1 + t_2 + \cdots + t_r, t_i \geqslant 0, r = 1, 2, \ldots\right\}$$

where

$$f_\alpha(x) = \log\left\{\Gamma(1+x)/(x+\alpha)^{x+\alpha}\right\}.$$

Clearly from the problem, $0 \leqslant \alpha \leqslant 1$.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING

COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Advanced Calculus. By Harold Edwards. Krieger, Huntington, New York, 1980. pp. xvii + 508. ISBN 0-89874-047-9.

A Formal Background to Mathematics: A Critical Approach to Elementary Analysis. By Robert Edwards. 2 vols. Springer-Verlag, New York, 1980. pp. ix + 1170. ISBN 0-387-90513-8.

SHELDON AXLER
Department of Mathematics, Michigan State University, East Lansing, MI 48824

Probably every mathematician has had many frustrating conversations with educated non-mathematicians who have no idea what mathematics is or what we do. It is easy to blame this state of affairs on the nature of mathematics, but perhaps we are also partly at fault. Most students who take calculus go no further in mathematics, so the content and style of calculus courses shape the perception of mathematics held by a large percentage of college graduates. Typical calculus courses and typical calculus textbooks, filled with vague and confusing definitions but short on mathematical reasoning and proofs, cannot convey the beauty we see in mathematics. We don't even require calculus students really to *understand* what a limit is, thus

removing from their education one of the key intellectual accomplishments of the nineteenth century.

Those students who still enjoy mathematics after taking calculus usually receive their first exposure to rigorous modern mathematics in a course called either elementary analysis or advanced calculus. Only now do we reveal enough about the structure of the real line so that the student can understand a proof of the Intermediate Value Theorem. The least upper bound axiom, or some other equivalent formulation of the completeness of the real line, is introduced, and finally we can prove that a continuous function on a closed interval attains its maximum. At this level we usually insist that students understand limits, uniform continuity, Riemann integration, etc. These courses often do much more than just correctly cover what could have been done right the first time. The calculus is extended to several variables, where the Fundamental Theorem of Calculus becomes Stokes' Theorem.

Harold Edwards has taken a bold and unusual approach to this subject in his book *Advanced Calculus*. He has relegated the study of limits, the real line, uniform continuity, etc., to the last chapter, which we would ordinarily expect to be devoted to Stokes' Theorem. He begins immediately with differential forms and gets to Stokes' Theorem very early in the book. The significance of this move is that many students think that the material at the end of the book is less important—after all, courses often don't reach the last chapter of the text. By bringing differential forms from the end-of-the-book ghetto into the bright light of chapter one, the author is shifting the focus of attention.

At first I believed that this approach couldn't be successful. I thought that the formalism of differential forms would be too formidable a structure for students at this level to master. Yet Harold Edwards has made it work. His definitions of differential forms and their integrals seem to be very natural. Plenty of examples are provided and everything is well motivated. For example, the Implicit Function Theorem is stated and proved first for affine maps.

Perhaps it appears that I have contradicted myself by making a statement about one of the key purposes of an advanced calculus course, noting that the book by Harold Edwards does not seem to fulfill it, and then saying that the book works. The crucial point is that students should be introduced to some real mathematics at this level, and what Harold Edwards presents is genuine mathematics, both in spirit and content. The student who has gone through this book will have some grasp of the *nature* of modern mathematics.

For most undergraduates, the insights offered by the standard advanced calculus course are probably more important than the benefits of taming differential forms, so I would be hesitant to recommend *Advanced Calculus* as the text for a typical undergraduate course. However, this book would be an exciting choice for an honors course or a course for graduate students (unfortunately many entering mathematics graduate students do not know this material). Any mathematician who has been away from differential forms for a while will find that Harold Edwards has provided a pleasant reintroduction. The most important feature of *Advanced Calculus* is that it is fun—it is fun to read the exercises, it is fun to read the comments printed in the margins, it is fun simply to pick a random spot in the book and begin reading. This is the way mathematics *should* be presented, with an excitement and liveliness that show why we are interested in the subject.

Robert Edwards (presumably no relation to the Harold Edwards of the first book under review) has written an entirely different type of book. He criticizes many texts for bad lapses from formally correct levels of rigor. These criticisms are often quite valid. It is all too easy to find first year calculus books in which a differential called dy is defined, but with a definition which is just nonsense. Such confusion, usually attributed to an attempt to make things easier for the student, probably hurts the weakest students most.

So Robert Edwards has produced a text which takes a formal approach to elementary analysis. *Elementary Analysis* (an abbreviation for the full title) intentionally contains few pictures or diagrams because the author believes that "diagrams have no place in a formal approach." The notation is so heavy that the reader will be distracted from the content. For example, having already defined the notion of a sequence converging to 0, we are told that a sequence u converges

to a real number c if $u - \underset{\sim}{c}'_N$ converges to 0. Of course, $\underset{\sim}{c}'_N$ denotes the obvious constant sequence, but is this notation really going to help anyone understand what's going on or prevent errors in logic?

As another example, having previously shown that $(1 + x/n)^n$ converges as $n \to \infty$ for each real number x, we are given the following definition of the exponential function:

$$\exp \equiv \operatorname{def}\{z : (\exists x)(\exists y)(x \in R \wedge y \in R \wedge (\text{the sequence } n \rightsquigarrow (1 + x/n)^n \text{ converges to the limit } y) \\ \wedge z = (x, y))\}.$$

And so it goes throughout the entire book.

To be fair, the author assumes that this material is not new to the reader, and that the reader merely wishes to review or to see the formal background which was previously hidden. However, it will be very difficult to use *Elementary Analysis* for review or reference. To understand the notation one must start at the beginning; there is so much cross-referencing that the chapters cannot be read independently. The constant irritation of petty formalism means that the book is no fun to pick up for a few minutes of browsing, and I can't imagine that anyone would enjoy reading the whole book.

Since the beginning of the nineteenth century mathematics has become significantly more formal and more rigorous. This development has given us deeper insight into the intricacies of analysis and has helped steer us away from serious errors. (Actually, it is remarkable how few errors Newton, Euler, etc. made while operating with what we now consider unacceptable sloppiness.) Vague and intuitive notions about infinitesimals have been replaced by precise ε-δ definitions. With nonstandard analysis we can now even remove infinitesimals from the realm of mysticism.

We certainly should not return to the carelessness of eighteenth-century mathematics. But Robert Edwards has inserted far too much formalism into *Elementary Analysis*. The purpose of formalism is to lead to clear and correct thinking, not to obscure the mathematical content. A super-excess of formality at an elementary level only leads to a dull and lifeless presentation. Too much mathematics, especially at the research paper level, is written with the meaning hidden from the reader. *Elementary Calculus* is explicitly aimed at students who will themselves become teachers of calculus. I hope that any future teachers who come across this book will not imitate its style to their students.

MISCELLANEA

69. In recent time there has been considerable unrest in the United States on matters relating to mathematics in secondary schools. Proceeding on the theory that the real aim is not, as formerly held, to "train the mind," but to teach only "things that are practical," so that the mathematical instruction can be "hitched up to life," some educators exert pressure to omit certain subjects formerly thought essential, such as the treatment of factoring and fractions in algebra, and to restrict problems in algebra and exercises in geometry to those that are practical. In some ways the modern tendency resembles the movement of the fifteenth and sixteenth centuries, led by the "Commercial School of Arithmeticians." As in the sixteenth century, so now, there is meant by "practical," not that which is *ultimately* practical, but that which is *immediately* practical.—F. Cajori, *A History of Elementary Mathematics with Hints on Methods of Teaching*, 2nd ed., New York, 1917, p. 303.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 4 April 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

OLD AND NEW APPROACHES TO EULER'S TRIGONOMETRIC EXPANSIONS

WOLFGANG WALTER

Universität Karlsruhe, Mathematisches Institut I, D-7500 Karlsruhe I Kaiserstrasse 12, Postfach 6380, West Germany

1. Introduction. One of the highlights in Euler's *Introductio in analysis infinitorum* of 1748 (later cited as *Introductio*) is the representation of the sine as an infinite product

$$\sin \pi x = \pi x \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{k^2}\right). \tag{1}$$

This result is the key to many other expansions and identities which fascinated mathematicians in the eighteenth century (and perhaps today). The logarithmic derivative of (1) generates the resolution of the cotangent into partial fractions

$$\pi \cot \pi x = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{2x}{x^2 - k^2} = \frac{1}{x} + \sum_{k=1}^{\infty} \left(\frac{1}{x+k} + \frac{1}{x-k}\right). \tag{2}$$

Among the remarkable corollaries which follow from (1) or (2) we mention only the Wallis product and the summation of the zeta function for all even positive integers, in particular

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}.$$

The problem of finding $\zeta(2)$ withstood the efforts of the best mathematicians, including the elder Bernoulli brothers, until it was finally solved by Euler in 1736; see [**1**].

We shall present here a derivation of (2) which is of surprising simplicity (considering the fact that the result has been known for some 250 years). The leading ideas were given by E. Mohr in a short note [**13**] which obviously remained unnoticed [in MR 14, p. 1080 and in Zentralbl. 50, p. 288 no word or comment is found]. A German textbook by Erwe [**6**] follows a slightly different route. This derivation of (2) seems to be within easy reach of undergraduate calculus. Before describing it, we shall comment on the history of the subject and on other ways of deducing (1) or (2).

2. Euler and Cauchy. Newton worked with power series as if they were polynomials. Euler carried this idea a bold step further. He generalized the factorization of polynomials to transcendental functions. In Chapter 9 of the *Introductio* he searches for the (real) quadratic factors of the polynomial $a^n - z^n$ (a, z real) and discovers that they are given by $a^2 + z^2 - 2az\cos(2k\pi/n)$ (§151). Since for Euler $e^x = (1 + (x/i)^i)$, where i is an infinitely large number (§155), the preceding result can be applied to

$$e^x - e^{-x} = \left(1 + \frac{x}{i}\right)^i - \left(1 - \frac{x}{i}\right)^i.$$

Using $\cos(2k\pi/i) = 1 - (2k^2\pi^2/i^2)$, the quadratic factors are found to be

$$\frac{4x^2}{i^2} + \frac{4k^2\pi^2}{i^2} - \frac{4k^2\pi^2x^2}{i^4}$$

The author received his doctorate from the University of Tübingen under the direction of E. Kamke. He is a Professor at Karlsruhe and has held visiting appointments at several universities in the United States. His research interests are mainly in differential equations, especially differential inequalities; his book *Differential and Integral Inequalities* appeared in 1970 in the Ergebnisse series.—*Editors*

or, after multiplication by $i^2/4k^2\pi^2$,

$$1+\frac{x^2}{k^2\pi^2}-\frac{x^2}{i^2}.$$

"Here, the member x^2/i^2 can be omitted, since it remains infinitely small, even when multiplied by i." In this way the representation

$$\frac{1}{2}(e^x-e^{-x})=x\left(1+\frac{x^2}{\pi^2}\right)\left(1+\frac{x^2}{4\pi^2}\right)\left(1+\frac{x^2}{9\pi^2}\right)\left(1+\frac{x^2}{16\pi^2}\right)\cdots$$

is obtained in §156. If x is an imaginary quantity, Euler writes in §158, then this formula changes over to a corresponding formula for the sine of a real arc. So much for Euler's derivation of (1). It gives a good insight in the way he deals with limiting processes.

Cauchy takes up the matter in the last pages of his celebrated *Cours-d'analyse* of 1821 [**3**]. His starting point is the factorization formula for the sine function (p. 462)

$$\sin z=m\sin\frac{z}{m}\cos\frac{z}{m}\left(1-\frac{\sin^2\frac{z}{m}}{\sin^2\frac{\pi}{m}}\right)\left(1-\frac{\sin^2\frac{z}{m}}{\sin^2\frac{2\pi}{m}}\right)\cdots\left(1-\frac{\sin^2\frac{z}{m}}{\sin^2\frac{(m-2)\pi}{2m}}\right),\tag{3}$$

valid for even integers $m\geqslant 2$. Letting $m\to\infty$ in (3), he obtains the sine product, using rather complicated calculations. A concise and elegant treatment along these lines is given by Bromwich [**2**, p. 213].

3. Other approaches. In a paper [**10**] written in 1868, H. Schröter gave a new and simpler deduction of (1) and (2) which is based on the trigonometric formulae for the half-angle. We describe his method in the case of (2). Since $2\cot x=\cot x/2-\tan x/2$, the function $\phi(x)=\cot\pi x$ satisfies the functional equation

$$\phi(x)=\frac{1}{2}\left[\phi\left(\frac{x}{2}\right)+\phi\left(\frac{x+1}{2}\right)\right].\tag{4}$$

By applying the formula twice to the expressions on the right, one obtains

$$\phi(x)=\frac{1}{4}\left[\phi\left(\frac{x}{4}\right)+\phi\left(\frac{x+1}{4}\right)+\phi\left(\frac{x+2}{4}\right)+\phi\left(\frac{x+3}{4}\right)\right]$$

and, after n such steps,

$$\phi(x)=\frac{1}{2^n}\sum_{k=0}^{2^n-1}\phi\left(\frac{x+k}{2^n}\right).\tag{5}$$

The inductive step for the general verification of (5) is immediate. It involves application of (4) to each term on the right,

$$\frac{1}{2^n}\phi\left(\frac{x+k}{2^n}\right)=\frac{1}{2^{n+1}}\left[\phi\left(\frac{x+k}{2^{n+1}}\right)+\phi\left(\frac{x+2^n+k}{2^{n+1}}\right)\right]$$

and summing over k. Equation (2) follows from (5) by letting $n\to\infty$. A well-written rigorous proof of this limiting process is given in K. Knopp's book on Infinite Series [**8**, pp. 205–206]. It is elementary, but still lengthy.

The theory of Fourier series offers a short proof of (2). It is found in many textbooks, e.g., in Courant and John [**5**, p. 602] and in Knopp [**8**, p. 377].

Several methods of obtaining expansions such as (1) or (2) are available using methods of complex analysis. In a memoir [**4**] written in 1851, Cauchy applies his theory of residues to the decomposition of transcendental functions into simple factors. His first example is the sine

product (1). While the verification of his reasoning presents some difficulties, the approach via residue theory has been simplified in various ways. Two such versions are given in the books by Titchmarsh [**11**, p. 113] and Levinson and Redheffer [**9**, p. 346], respectively.

The product theory of entire analytic functions, which Weierstrass developed in Chapter 12 of his *Vorlesungen über die Theorie der elliptischen Funktionen* [**12**] (written before 1863 and published in 1915), leads directly to the sine product (1). By applying his product theorem to the sine function, Weierstrass finds (p. 112) that

$$\sin \pi u = e^{g(u)} u \prod_{k=-\infty}^{\infty} \left(1 - \frac{u}{k}\right) e^{u/k} = e^{g(u)} u \prod_{k=1}^{\infty} \left(1 - \frac{u^2}{k^2}\right).$$

By differentiating $\log \sin u$ twice and letting $\operatorname{Im} u \to \infty$, he obtains the relations $g'' = 0$ and finally $e^{g(u)} = \pi$. The determination of $g(u)$ can be shortened by using stronger results from the value distribution theory developed by Hadamard.

4. Investigation of the series in (2). We now turn to the proof of equation (2) along the lines given by Mohr. Let f be the function defined by the right-hand side of (2),

$$f(x) = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{2x}{x^2 - k^2} = \frac{1}{x} + \sum_{k=1}^{\infty} \left(\frac{1}{x+k} + \frac{1}{x-k}\right).$$

Let N be a large positive integer and $|x| \leqslant N$. The inequality

$$\left|\frac{2x}{x^2 - k^2}\right| \leqslant \frac{4N}{k^2} \quad \text{for } k \geqslant 2N$$

shows that the series is uniformly convergent in bounded subsets of $\mathbb{R} \backslash \mathbb{Z}$ and that f is a continuous function in $\mathbb{R} \backslash \mathbb{Z}$. One verifies by inspection that f is an odd function. Now let

$$S_m^n(x) = \sum_{k=m}^{n} \frac{1}{x+k}.$$

Then $f(x) = \lim_{n \to \infty} S_{-n}^n(x)$ and, more generally,

$$f(x) = \lim_{n \to \infty} S_{-n+q}^{n+p}(x),$$

when the numbers p, q are taken from the set $\{0, 1, -1\}$. In the identity

$$S_{-n}^n(x+1) = S_{-n+1}^{n+1}(x)$$

the left-hand side converges to $f(x+1)$ and the right-hand side converges to $f(x)$ as $n \to \infty$. Hence $f(x) = f(x+1)$. In a similar way, the equation

$$\frac{1}{\frac{x}{2} + k} + \frac{1}{\frac{x+1}{2} + k} = \frac{2}{x + 2k} + \frac{2}{x + 2k + 1}$$

leads to

$$S_{-n}^n\left(\frac{x}{2}\right) + S_{-n}^n\left(\frac{x+1}{2}\right) = 2 S_{-2n}^{2n+1}(x).$$

Letting $n \to \infty$, it follows that

$$f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right) = 2f(x).$$

Summing up, we have shown that the function f shares the following properties with the function $\pi \cot \pi x$: For $x \in \mathbb{R} \backslash \mathbb{Z}$, it is continuous, odd, 1-periodic, and a solution of equation (4).

5. Conclusion of Proof. The function ϕ defined by

$$\phi(x) := \pi \cot \pi x - f(x)$$

has the four properties stated at the end of Section 4. Next we show that $\phi(x) \to 0$ as $x \to 0$.

For $|x| \leq \frac{1}{2}$,

$$\sum_{k=1}^{\infty} \left| \frac{2x}{x^2 - k^2} \right| \leq 4|x| \sum_{k=1}^{\infty} \frac{1}{k^2} \to 0 \quad \text{as } x \to 0.$$

Therefore, we have to show that $\lim_{x\to 0}(\pi \cot \pi x - (1/x)) = 0$ or, equivalently, $\lim_{x\to 0}(\cot x - (1/x)) = 0$. Using power series, we get

$$\cot x = \frac{\cos x}{\sin x} = \frac{1}{x} \cdot \frac{1 - \dfrac{x^2}{2} + \dfrac{x^4}{24} - + \cdots}{1 - \dfrac{x^2}{6} + \dfrac{x^4}{120} - + \cdots} = \frac{1}{x}\left(1 + c_1 x^2 + c_2 x^4 + \cdots\right),$$

from which the desired result follows.

Another argument which uses, instead of power series expansions, only the inequalities

$$0 < \sin x < x < \tan x \quad \text{for small positive } x,$$

is just as easy. These inequalities imply

$$0 < \frac{1}{x} - \cot x < \frac{1}{\sin x} - \cot x = \frac{1 - \sqrt{1 - \sin^2 x}}{\sin x} = \frac{\sin x}{1 + \sqrt{1 - \sin^2 x}} < x.$$

Here we used the identity $(1 - \sqrt{1-a})(1 + \sqrt{1-a}) = a$. Hence the function $\cot x - (1/x)$ tends to zero as $x \to +0$. Since this function is odd, the limit as $x \to -0$ vanishes, too.

Because of periodicity, $\phi(x) \to 0$ as $x \to k$, where k is any integer. By putting $\phi(k) = 0$ for $k \in \mathbb{Z}$, the function ϕ becomes a continuous function in $\mathbb{R}$.

We are now ready to show that ϕ vanishes identically. Assume to the contrary, that R is a (large) constant and that

$$M := \max\{|\phi(x)| : |x| \leq R\} > 0.$$

There exists ξ, $|\xi| \leq R$, such that $|\phi(\xi)| = M$. Since $|\xi/2|, |(\xi+1)/2| \leq R$, equation (4) yields

$$|\phi(\xi)| = M \leq \frac{1}{2}\left\{\left|\phi\left(\frac{\xi}{2}\right)\right| + \left|\phi\left(\frac{\xi+1}{2}\right)\right|\right\} \leq \frac{1}{2}(M + M) = M.$$

Here we must have $|\phi(\xi/2)| = M$, because $|\phi(\xi/2)| < M$ would lead to a strict inequality $M < M$. A repetition of this argument gives $|\phi(\xi/4)| = |\phi(\xi/8)| = \cdots = M > 0$. On the other hand, $\phi(\xi) \to 0$ as $\xi \to 0$. This contradiction completes the proof.

In the last step we followed Erwe, while Mohr has given an alternative proof, which is of interest in itself and is therefore stated as a theorem.

6. Theorem. If ϕ is Riemann integrable in [0, 1] and if ϕ satisfies the functional equation

$$\phi(x) = \frac{1}{2}\left[\phi\left(\frac{x}{2}\right) + \phi\left(\frac{x+1}{2}\right)\right] \quad \text{for } 0 < x < 1, \tag{4}$$

then

$$\phi(x) = \text{const.} = \int_0^1 \phi(t)\, dt \quad \text{for } 0 < x < 1.$$

The proof consists of recognizing that the right-hand side of equation (5),

$$\frac{1}{m}\sum_{k=0}^{m-1}\phi\left(\frac{x+k}{m}\right) \quad \text{with } m=2^n,$$

is, for any $x \in (0,1)$, a Riemann sum for the integral $\int_0^1\phi(t)\,dt$ corresponding to partitioning the interval $[0,1]$ into m subintervals of length $1/m$ (for example, the choice $x = 1/2$ corresponds to taking function values at the midpoints of the respective subintervals). Since ϕ satisfies (5) and the Riemann sums converge to the integral in question, the assertion follows.

This theorem clearly implies that our function ϕ (considered earlier) vanishes in $[0,1]$ and, by periodicity, in $\mathbb{R}$.

7. Complex variables. Once equation (2) is established for real x, it carries over to complex values of x by analytic continuation. But analytic continuation can be avoided, since the above proof carries over to complex variables. It follows from Section 4 that f is a continuous, 1-periodic solution of (4) in $\mathbb{C}\backslash\mathbb{Z}$, and from Section 5 that ϕ is continuous in $\mathbb{C}$ and finally that $\phi \equiv 0$. Alternatively, a complex version of Theorem 6 can be used in the last step. If ϕ is continuous in the strip $0 \leqslant \operatorname{Re} x \leqslant 1$ and if ϕ satisfies Equation (4) in the open strip $0 < \operatorname{Re} x < 1$, then the Riemann sums considered above converge again to the integral given in Theorem 6, i.e., ϕ is constant.

8. The Passage from (2) to (1). Equation (2) is derived from (1) by taking logarithmic derivatives on both sides. The reverse procedure of inferring (1) from (2) presents no difficulty, provided some knowledge about the convergence and the derivative of infinite (series and) products is brought in; see, e.g., Knopp [**8**, p. 383–384].

Because of the elementary character of our presentation it is worthwhile to give an independent proof which uses no such premises. We define polynomials P_m by

$$P_m(x) := \pi x\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\cdots\left(1-\frac{x^2}{m^2}\right)$$

and show that $\lim P_m(x) = \sin \pi x$. This is, so to speak, our definition of Equation (1).

The only possible zeros or points of discontinuity of the function

$$Q_m(x) = \frac{P_m(x)}{\sin \pi x}$$

are integers. Since $(1-(x^2/k^2))/\sin \pi x$ tends to a nonzero limit as $x \to k \in \mathbb{Z}$, the function Q_m is, when properly defined for integer values, a continuous and positive function in the interval $0 \leqslant x \leqslant m$. We have

$$\frac{d}{dx}\log Q_m(x) = S_{-m}^{m}(x) - \pi\cot \pi x = \sum_{k=m+1}^{\infty}\frac{2x}{k^2-x^2} \equiv R_m(x).$$

For positive x and $m > 2x$,

$$0 \leqslant R_m(x) \leqslant 4x\sum_{m+1}^{\infty}\frac{1}{k^2} < 4x\sum_{m+1}^{\infty}\frac{1}{k(k-1)} = \frac{4x}{m}.$$

Integration between 0 and $x > 0$ yields

$$0 \leqslant \log Q_m(x) \leqslant \frac{2x^2}{m}$$

or

$$1 \leqslant Q_m(x) \leqslant e^{2x^2/m}.$$

Observing that $P_m(x) - \sin \pi x = [Q_m(x) - 1]\sin \pi x$ and $e^t - 1 < 2t$ for $0 < t < 1$, we obtain

the final estimate

$$|P_m(x) - \sin \pi x| \leq |\sin \pi x| \frac{4x^2}{m} \quad \text{for } m > 2x^2 + 1.$$

How about complex values of x? This problem is left to the reader.

References

1. R. Ayoub, Euler and the Zeta function, this MONTHLY, 81 (1974) 1067–1085.
2. T. J. I'a. Bromwich, An Introduction to the Theory of Infinite Series, 2nd ed., London, 1926.
3. A. Cauchy, Cours d'analyse, 1re partie, Analyse algébrique, 1821 [= Oeuvres complètes, II^{e} Série, Tome III].
4. A. Cauchy, Mémoire sur l'application du calcul des résidues à la décomposition des fonctions transcendentes en facteurs simples, C. R., 32 (1851) 267 [= Oeuvres complètes, 1re Série, Tome XI, p. 314–424].
5. R. Courant and F. John, Introduction to Calculus and Analysis, vol. I, Interscience, 1965.
6. F. Erwe, Differential- und Integralrechnung I, BI-Hochschultaschenbuch, no. 30/30a, Mannheim, 1970.
7. L. Euler, Introductio in analysin infinitorum, Lausanne 1748 [= Opera omnia, Series prima, Volumen III].
8. K. Knopp, Theory and Application of Infinite Series, Blackie, London and Glasgow 1928; reprinted, Hafner, New York, 1948.
9. N. Levinson, and R. M. Redheffer, Complex Variables, Holden-Day, 1970.
10. H. Schröter, Ableitung der Partialbruch- und Produktentwicklungen für die trigonometrischen Funktionen, Z. Math. Phys., 13 (1868) 254–259.
11. E. C. Titchmarsh, The Theory of Functions, 2nd ed., Oxford University Press, 1939.
12. K. Weierstrass, Mathematische Werke, 5 (1915) 103–121.
13. E. Mohr, Elementarer Beweis fur die partialbruchzerlegung des Cotangens, Z. Angew. Math. Mech., 33 (1953) 247–248.

HOW A DIGITAL COMPUTER CAN TELL WHETHER A LINE IS STRAIGHT

AZRIEL ROSENFELD

Computer Science Department, University of Maryland, College Park, MD 20742

CHUL E. KIM

Computer Science Department, Washington State University, Pullman, WA 99164

1. Introduction. *Image processing* and *pattern recognition* [1] are often concerned with classifying shapes or patterns that appear in pictures, and the classification is often based on geometrical properties of the patterns. For example, in a picture of a nuclear bubble chamber, we may want to classify the particle tracks as being straight line segments, circular arcs, etc., in order to identify the particles that gave rise to these tracks. As another example, in a photomicrograph of a blood smear, we may want to determine whether the nucleus of a white blood cell is convex or has concavities in order to identify which type of cell it is.

What makes such tasks nontrivial is that computers can only deal with pictures that have been "digitized," i.e., converted into arrays of lattice points, and it is not always obvious how to

---

Chul E. Kim received his Ph.D. in Computer Science from the University of Minnesota in 1975. He is currently an Associate Professor of Computer Science at Washington State University.

Azriel Rosenfeld received the Ph.D. in mathematics from Columbia University in 1957. After ten years in the defense electronics industry, he joined the University of Maryland in 1964, where he is Research Professor of Computer Science. He edits the journal *Computer Graphics and Image Processing*, and is president of the consulting firm ImTech, Inc. He has published 13 books and over 250 papers, most of them dealing with the computer analysis of pictorial information. He is currently President of the International Association for Pattern Recognition.

recognize that a set of lattice points must have arisen from a real pattern that has a given geometric property. For example, how do we characterize sets of lattice points that are the digitizations of real straight line segments? This and some related questions will be discussed in this paper. In order to treat them, we must first define more precisely what we mean by "digitization," and introduce some basic "digital picture" terminology. As we shall see, the results depend strongly on the definitions of digitization that we use.

2. Digitization of bounded subsets. Let S be a bounded subset of the plane. For purposes of computer analysis, it is customary to represent S by a finite set of lattice points, i.e., points with integer coordinates. This set $\hat{S}$ is called the *digital image* of S, and the mapping that takes S into $\hat{S}$ is called *digitization*.

$\hat{S}$ can be defined in a number of ways; we list several of them here:

(a) $\hat{S}$ is the set of lattice points contained in S; this is called the *subset digitization* of S.

(b) $\hat{S}$ is the set of lattice points such that S comes closer than city block distance $\frac{1}{2}$ to them—i.e., $\{(i, j) \mid \exists(x, y) \in S\colon \max(|x-i|, |y-j|) < \frac{1}{2}\}$. This is called the *open cell digitization* of S. (If we imagine an open unit square ["cell"] P° centered at each lattice point P, we have $P \in \hat{S}$ iff $S \cap P^\circ \neq \varnothing$.)

(b′) Analogous to (b), using half-open cells P^*, e.g., $i - \frac{1}{2} \leq x < i + \frac{1}{2}$, $j - \frac{1}{2} \leq y < j + \frac{1}{2}$.

(b″) Analogous to (b), using closed cells $\bar{P}$.

Note that by definitions (a-b), a nonempty set S can have an empty digitization. In the course of this paper we will discover other advantages and disadvantages of the various definitions. A simple example of digitization is given in Fig. 1.

(a) (b) (c) (d)

FIG. 1. Simple example of digitization according to various definitions.
(a) Lattice points and set S (black blob). The subset digitization of S is empty.
(b) Open, (c) half-open, and (d) closed cell digitizations of S.

A set T of lattice points is called 8-*connected* if for all P, Q in T there exists a finite sequence $P = P_0, P_1, \ldots, P_n = Q$ of points of T such that P_i is a horizontal, vertical, or diagonal neighbor (for brevity: an 8-*neighbor*) of P_{i-1}, $1 \leq i \leq n$. If only horizontal and vertical neighbors ("4-neighbors") are allowed, we call T 4-*connected*.

PROPOSITION 1. *If S is arcwise connected, then by definition* (a) *or* (b), *$\hat{S}$ need not be* 8-*connected; by definition* (b′), *it must be* 8-*connected; and by definition* (b″), *it must be* 4-*connected.*||

Further properties of 4- and 8-connectedness are treated in [2], [3].

3. Digitization of arcs. None of the definitions of digitization given in Section 2 is entirely satisfactory if S is an arc. As we traverse an arc A, we would like to define a sequence of lattice points belonging to the digitization of A, and we would also like the digitization of an arc to be connected. The connectedness requirement immediately rules out the subset and open cell definitions (a-b); while if we use the closed cell definition (b″), the lattice points of $\hat{A}$ do not occur in a simple sequence; when A leaves a cell through one of its corners, three new lattice points (at the centers of the other cells sharing that corner) appear simultaneously on $\hat{A}$. This leaves only the half-open cell definition (b′), for which the lattice points of $\hat{A}$ do in fact occur in sequence as A is traversed. Each of these points is an 8-neighbor of the preceding one, so that $\hat{A}$ is determined by specifying a starting point and a sequence of moves from neighbor to neighbor [4].

This approach provides a compact way of specifying $\hat{A}$, but it is somewhat wasteful in the sense that diagonal moves occur with zero probability; when an arc leaves a cell, it almost certainly does so along a side, not at a corner. For this reason, a different definition of digitization has historically been used for arcs, which we may call *grid digitization*. Imagine the lattice points joined by a grid of lines; thus as we traverse A, we cross a succession of grid lines. (Note that $\hat{A}$ can be empty if A never crosses a grid line; but then A always stays inside a single cell.) Whenever we cross a grid line, the lattice point (= grid line intersection) closest to the crossing point becomes a point of $\hat{A}$. If we cross halfway between two lattice points, we resolve the tie by using, e.g., the lattice point that lies to the right of A (in the sense that we are traversing it)*. This grid digitization evidently defines a sequence of lattice points in $\hat{A}$ as A is traversed, each an 8-neighbor of the preceding; but it is easily seen that diagonal neighbors now have nonzero probability. A further advantage of grid digitization over cell digitization will become apparent in the next section. Figure 2 gives an example of arc digitization.

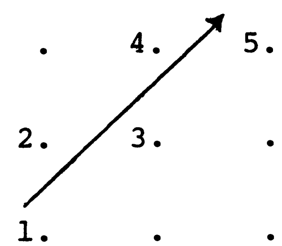


FIG. 2. Simple example of arc digitization. If we use cell digitizaiion, or grid digitization but resolve ties by rounding, the digitization of the line contains the sequence of lattice points $1, 2, 3, 4, 5, \ldots$, and is a 4-arc; but if we resolve ties by using the lattice point on the right of the line, it contains only $1, 3, 5, \ldots$, and is an 8-arc.

4. Digital arcs. The finite set of lattice points B is called a *digital arc* if

(a) B is connected.

(b) All but two points of B have exactly two neighbors in B.

(c) Two points of B, called the *endpoints*, have exactly one neighbor in B.

Note that this is two definitions in one, depending on whether we use the 4- or 8-definition for "neighbor" and "connected."

PROPOSITION 2. *If B is a digital arc, and we use the subset, open cell, or grid definition of digitization, then there exists an arc A such that $B = \hat{A}$.*

Proof. If we start from one of the endpoints, go to its neighbor, then go to the other neighbor of that neighbor (if any), and repeat the process, we can keep on until we reach a point that has no other neighbor, which must be the other endpoint. It is not hard to see that since B is connected, the sequence of points defined in this way is all of B. (If a point in the sequence were connected to a point not in the sequence, some point in the sequence would have to have a third neighbor.) The polygonal arc A joining this succession of points then evidently has B as its digitization by the three definitions mentioned. Note that the Proposition is not true for the other two definitions—e.g., the digital 8-arc $\{(0,0), (1,1)\}$ is not the closed cell digitization of any arc.||

Unfortunately, if A is an arc, $\hat{A}$ need not be a digital arc, since $\hat{A}$ may touch itself if A passes sufficiently close to itself. However, we can prove

PROPOSITION 3. *If A is a straight line segment, and we use the grid definition of digitization, then $\hat{A}$ is a digital 8-arc.*

*Alternatively, we could resolve ties by rounding, but as we shall see, the method defined here is preferable.

Proof. As we move along A, we visit the points of $\hat{A}$ in succession. It is not hard to see that the successive points of $\hat{A}$ (if distinct) are 8-neighbors, and that two points of $\hat{A}$ cannot be 8-neighbors unless they are successive.||

Proposition 3 does not hold if we use the subset or cell digitizations, or even if we use the grid method but resolve ties by rounding. For the subset or open cell method, $\hat{A}$ can evidently be empty; and for the closed cell method, the line $x = i \pm \frac{1}{2}$ or $y = j \pm \frac{1}{2}$ has a double-thickness digitization. Even for the half-open cell method, let A be the line through $(\frac{1}{2}, \frac{1}{2})$ with slope $-45°$; then $\hat{A}$ defined by the half-open cell method has a digitization that is a 4-arc, not an 8-arc, since it contains the lattice points $\ldots, (1,0), (1,1), (0,1), (0,2), (-1,2), \ldots$. Similarly, let A be the line through $(\frac{1}{2}, 0)$ with slope $45°$, and let $\hat{A}$ be defined by the grid method but with ties resolved by rounding down; then $\hat{A}$ is a 4-arc but not an 8-arc, since it contains the lattice points $(0,0), (1,0), (1,1), (2,1), \ldots$. (The same example works if we round up rather than down; and if we round up in one coordinate and down in the other, we can give an analogous example using a line of slope $-45°$.)

We are now ready to consider the question posed in the title of this paper: Given a digital arc, how can we tell whether it is the digitization of a straight line segment? Note that any digital arc is always the digitization of things that are not straight line segments, but we want to know when it is also the digitization of a straight line segment.

5. Straight digital arcs. A digital 8-arc B will be called *straight* if there exists a straight line segment A such that $\hat{A} = B$ (using grid digitization).

THEOREM 4. *The following properties of the digital arc B are equivalent*:
(a) *B is straight.*
(b) *There exists no triple of collinear lattice points P, Q, R (with Q between P and R) such that P, R are in B but Q is not.*
(c) *For any lattice points P, R of B, and any point (x, y) on the line segment $\overline{PR}$, there exists a lattice point (i, j) of B such that* $\max(|x - i|, |y - j|) < 1$.||

It is not hard to show that if B is straight, it has properties (b-c). Conversely, we can easily show that if B has property (b) or (c), its sequence of moves from neighbor to neighbor can involve at most two directions, which can only differ by $45°$, and that at least one of these directions has only isolated occurrences in the sequence; thus the sequence consists of runs in a given direction, separated by single moves in an adjacent direction. Now if property (b) or (c) holds for B, it also holds for the digital arc B' obtained by deleting the last point of B; hence by induction, B' is straight, say $B' = \hat{A}'$; and whether the last point P of B extends a run or starts a new run, one can find an A' such that P is on the digitization of an extension of A'. For the details of a proof that (a) and (c) are equivalent, see [**5**].

A set of lattice points for which (b) holds will be said to have the *collinearity property*, and a set for which (c) holds will be said to have the *chord property*. At first glance, (b) and (c) seem tedious to verify; but in fact, they need only be checked for pairs P, R of lattice points of B that are run ends, and (as regards (c)) for points (x, y) that have the same coordinate as a run end, since run ends are extrema of the deviation of the digital arc from straightness. Theorem 4 is not true for other definitions of digitization; $\{(0,0), (1,1)\}$ and $\{(0,1), (1,0)\}$ are digital 8-arcs, and evidently have properties (b-c), but as we have already seen, by the other definitions they are not both digitizations of straight line segments.

Other aspects of digital straightness discussed in the literature include grammars for the move sequences in a digitized straight line [**6**], [**7**], integer equations for the lattice points on such a line [**8**], and accuracy in determining a line's position from its digitization [**9**], as well as properties related to those of Theorem 4 [**10**], [**11**], [**12**].

6. Digital convexity. The conditions of Theorem 4 turn out to be of interest for other reasons;

in fact, they are precisely the conditions for a set of lattice points to be the digitization of a convex set, if we use the subset definition of digitization. To begin with, it can be shown that

THEOREM 5. *The following properties of a finite set T of lattice points are equivalent*:

(a) *T has the collinearity property.*

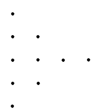
(b) *T has the chord property.*

(c) *The convex hull of T contains no lattice point in the complement of T.*||

T is called *digitally convex* if there exists a convex set S such that $\hat{S} = T$.

THEOREM 6. *T is digitally convex* (*using the subset definition of digitization*) *iff it has the properties of Theorem 5.*||

For other definitions of digitization, Theorem 6 does not hold. If we use the cell definitions, the conditions of Theorem 5 are necessary but not sufficient. As an example, let T be



Then T has the properties of Theorem 5, but is not digitally convex by any of the cell definitions.

THEOREM 7. *T is digitally convex* (*subset definition*) *iff for any two lattice points P, Q of T there exists a straight digital 8-arc B such that $P, Q \in B \subseteq T$.*||

Proofs of these results (in whole or in part), and algorithms for verifying convexity and straightness, for various definitions of digitization, will appear in [**13**–**16**]; see also [**17**] on digital polygons. Earlier work on digital convexity deals with various partial convexity criteria, notably involving minimum-perimeter polygons that have the given digitization [**18**–**21**] ([**19**] also discusses digitization in general), as well as with methods of concavity detection and filling [**22**–**25**].

It should be noted that the situation is more complex in three dimensions [**26**]. For example, it can be shown that when we use a method analogous to open cell digitization, the chord property is sufficient but not necessary for a set of lattice points in three dimensions to be the digitization of a convex object. Three-dimensional digital geometry is a subject of rapidly growing interest with the increasing need to process three-dimensional data arrays, e.g., as obtained by computed tomography.

7. Concluding remarks. Determining whether a sequence of lattice points could be the digitization of a straight line segment, or a set of lattice points the digitization of a convex object, is of practical interest in digital image processing and pattern recognition. These problems turn out to have neat solutions for some definitions of digitization, but not for others. Thus the method of digitization used to represent planar subsets in a computer can have unexpected implications with respect to determining geometric properties of the subsets.

References

1. A. Rosenfeld and A. C. Kak, Digital Picture Processing, Academic Press, New York, 1976.

2. A. Rosenfeld, Picture Languages, Academic Press, New York, 1979, Chapter 2.

3. ______, Digital topology, this MONTHLY, 86, 1979, 621–630.

4. H. Freeman, Computer processing of line drawing images, Comput. Surveys, 6 (1974) 57–97.

5. A. Rosenfeld, Digital straight line segments, IEEE Trans. Computers, 23 (1974) 1264–1269.

6. R. Brons, Linguistic methods for the description of a straight line on a grid, Computer Graphics Image Processing, 3 (1974) 48–62.

7. J. Rothstein and C. Weiman, Parallel and sequential specification of a context sensitive language for straight lines on grids, Computer Graphics Image Processing, 5 (1976) 106–124.

8. G. Bongiovanni, F. Luccio, and A. Zorat, The discrete equation of the straight line, IEEE Trans. Computers, 24 (1975) 310–313.

9. H. Klaasman, Some aspects of the accuracy of the approximated position of a straight line on a square grid, Computer Graphics Image Processing, 4 (1975) 225–235.

10. C. Arcelli and A. Massarotti, Regular arcs in digital contours, Computer Graphics Image Processing, 4 (1975) 339–360.

11. M. Gaafar, Convexity verification, block-chords and digital straight lines, Computer Graphics Image Processing, 6 (1977) 361–370.

12. C. Arcelli and A. Massarotti, On the parallel generation of straight digital lines, Computer Graphics Image Processing, 7 (1978) 67–83.

13. C. E. Kim, On the cellular convexity of complexes, IEEE Trans. Pattern Analysis Machine Intelligence, in press.

14. C. E. Kim and A. Rosenfeld, Digital straight lines and convexity of digital regions, ibid.

15. C. E. Kim, On cellular straight line segments, Computer Graphics Image Processing, in press.

16. C. E. Kim and J. Sklansky, Digital and cellular convexity, submitted to Pattern Recognition.

17. C. E. Kim, Digital convexity, straightness, and convex polygons, TR-1055, Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD, May 1981.

18. J. Sklansky, Recognition of convex blobs, Pattern Recognition, 2 (1970) 3–10.

19. G. U. Montanari, On limit properties in digitization schemes, J. Assoc. Comp. Mach., 17 (1970) 348–360.

20. L. Hodes, Discrete approximation of continuous convex blobs, SIAM J. Appl. Math., 19 (1970) 477–485.

21. J. Sklansky, R. L. Chazin, and B. J. Hansen, Minimum-perimeter polygons of digitized silhouettes, IEEE Trans. Computers, 21 (1972) 1233–1239.

22. J. Sklansky, Measuring concavity on a rectangular mosaic, IEEE Trans. Computers, 21 (1972) 260–268.

23. C. Arcelli and L. Cordella, Concavity point detection by iterative arrays, Computer Graphics Image Processing, 3 (1974) 34–47.

24. J. Sklansky, On filling cellular concavities, Computer Graphics Image Processing, 4 (1975) 236–247.

25. J. Sklansky, L. P. Cordella, and S. Levialdi, Parallel detection of concavities in cellular blobs, IEEE Trans. Computers, 25 (1976) 187–196.

26. C. E. Kim, Convex digital solids, submitted to IEEE Trans. Pattern Analysis Machine Intelligence.

NEW CONCEPTS IN SEEDING KNOCKOUT TOURNAMENTS

F. K. HWANG
Bell Laboratories, Murray Hill, NJ 07974

1. Introduction. A *knockout tournament* with $n = 2^r$ players is usually defined (see [**2**], [**3**]) in the following way: The tournament consists of r rounds of matches where in the ith round, the 2^{r+1-i} survivors are paired off for 2^{r-i} matches, with the winners moving on to the next round and the losers dropping out from the tournament. The one player left at the end of the rth round is the tournament *champion*. Therefore the two essential features of a knockout tournament are that every player except the champion loses exactly one match and the champion wins exactly r matches while losing none. Knockout tournaments have been defined in a more general way [**1**] by not requiring the champion to win exactly r matches (then the tournament described here would have been called a "balanced" knockout tournament). However, in this paper we hold to the definition given, and we comment that most knockout tournaments in the real world are of the "balanced" type or a slight modification of it in which some players may have a bye round (to be clarified later).

Let $P = \{p_{ij}\}$ be a set of probabilities satisfying $p_{ij} + p_{ji} = 1$ where p_{ij} denotes the probability

In answer to the editors' request for a condensed autobiography the author writes as follows.

I received my Ph.D. from North Carolina State University in 1968. I have worked at Bell Laboratories since 1967, where I am now with the department of discrete mathematics. I have been on leave three times: at National Tsing-Hua University (Taiwan), 1970; at Telecommunications Laboratories (Taiwan), 1976; and at Academia Sinica (Beijing), 1980. My main research interest is in combinatorics. I have been on the national bridge team of Taiwan five times, including the 1969 Bermuda Bowl runner-up team and the 1972 World Bridge Olympiad fifth-place team.

that player x_i beats player x_j in a match. P is usually known as the *preference scheme* for the involved players. David [2] defined P to be *strong stochastically transitive* (hereafter abbreviated as *SST*) if for every i and j, $p_{ik} > p_{jk}$ for some k implies $p_{ik} \geqslant p_{jk}$ for every k. Thus if P is SST, then the players can be ranked according to their playing strength in a strong sense. Namely, player x_i is ranked above player x_j (or we say that x_i is *stronger* than x_j) if x_i has a better chance of beating x_j than vice versa; furthermore, in this case x_i has a better chance of beating any one else than x_j does. In this paper we are only concerned with preference schemes which are SST, and we assume that the players are ranked in the order $x_1, x_2, \ldots, x_n$, from highest to lowest.

If the pairings are done randomly, then it can happen that good players knock each other out in early rounds while weaker players survive until much later rounds. To prevent this, "seeding" is often used for knockout tournaments. Justifiably or not, one main purpose of seeding is to bias the tournament in favor of stronger players by pairing them with weaker players in early rounds. Even though this purpose is debatable in theory, it has not encountered much resistance in practice. However, no one expects a good or "fair" seeding to bias in favor of weaker players. We embody this principle by calling a seeding *monotone* if x_i stronger than x_j implies that x_i has at least as large a probability of winning the tournament as x_j under the given seeding.

The purpose of this paper is to show that the traditional methods of seeding, including some very popular ones which are in use in many important tournaments, are not necessarily monotone. We then suggest a new seeding method which is monotone for all preference schemes which are SST. We also show how to implement this seeding method when the rank-order of the players is not unanimously agreed upon.

2. The traditional methods of seeding. The traditional methods of seeding can be best described by introducing the notion of a binary tree. A *binary tree* can be recursively defined by specifying that it consists of either a *root* or a root linked with two nonempty subtrees which are also binary trees. The links are considered directed from the root to the subtrees. A root of a subtree is called an *internal node* or a *terminal node* depending on whether that root is linked to any nonempty subtrees or not. A binary tree is *balanced* if the number of links between a terminal node and the root of the tree is constant. Fig. 1 shows a balanced binary tree with eight terminal nodes.

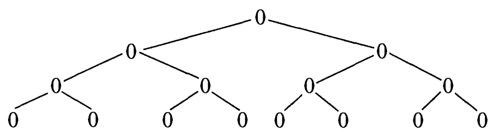


FIG. 1. A Balanced Binary Tree.

A knockout tournament with $n = 2^r$ players can be represented by a balanced binary tree with n terminal nodes. A traditional seeding is a labeling of the terminal nodes each by a distinct player. Two labeled nodes linked to the same internal node constitute a match with the winner then labeling the above-mentioned internal node. This labeling process goes on until the root of the binary tree is labeled by the tournament champion. In this representation, a seeding on n players $x_1, x_2, \ldots, x_n$, can be described by a permutation $(\pi_1, \pi_2, \ldots, \pi_n)$ of the set $\{1, 2, \ldots, n\}$ with the understanding that player x_{π_j} labels the jth terminal node, counting from left to right.

We now describe a seeding method S which can be considered as the *standard* method for traditional seeding. We define S recursively: For $r = 1$, the seeding is $1, 2$. Suppose $k_1, k_2, \ldots, k_{2^{r-1}}$ is the seeding for 2^{r-1} players. Then the seeding for $n = 2^r$ players is $k_1, n + 1 - k_1, k_2, n + 1 - k_2, \ldots, k_{2^{r-1}}, n + 1 - k_{2^{r-1}}$. For example, the seedings S for four and eight players are $1, 4, 2, 3$ and $1, 8, 4, 5, 2, 7, 3, 6$, respectively.

The method S also works when n is not a power of two, say, $n = 2^\alpha + \beta, 0 < \beta < 2^\alpha$. $2^\alpha - \beta$ dummy players are added to increase the number of players to $n' = 2^{\alpha+1}$. These dummy players

are all ranked below any regular player and they lose to regular players with certainty. When S is applied to n', all the dummy players disappear after the first round.

The method S is intuitively appealing and is widely used in many real situations. The method seems to give as much advantage to stronger players as possible and any improvement in that respect is hard to find. However, we now show that there exist SST preference schemes for which the standard seeding is not monotone.

Consider the preference scheme shown in Fig. 2.

$i \backslash j$	2	3	4	5	6	7	8
1	p	p	p	p	1	1	1
2		p	p	p	1	1	1
3			p	p	p	1	1
4				p	p	1	1
5					p	1	1
6						1	1
7							1

where $p = .5 + \varepsilon$

FIG. 2. A Preference Scheme.

It is easily verified that under S and assuming ε small, player x_1 has probability $1/4$ but player x_2 has probability $3/8$, of winning the tournament. What causes the anomaly is the fact that player x_2 has a fifty percent chance of playing against player x_6, an almost certain win, in the second round. This example can be easily modified to an example for six players by treating x_7 and x_8 as dummy players. It is also obvious that similar examples can be constructed for larger n. Therefore we see that the method S may even hurt the top-ranked player. Furthermore, it is hard to find any other reasonable criterion (such as minimizing the expected ranks of the winners of the ith round) which the method S is guaranteed to optimize for all SST preference schemes.

3. A new seeding method. We now describe a new seeding method M for $n = 2^r$ players (if n is not a power of two, we use the same "addition of dummy players" trick as described for S). The main difference between M and S is that seeding is done in every round under M and that seeding in the kth round depends only on the relative ranks of the players in that round, but not on the seedings of previous rounds. Let y_i denote the ith strongest player among the players of the kth round. Then M specifies that during round k, y_i is paired off with y_j where $j = 2^{r+1-k} - i + 1$. We will prove that M is monotone with respect to any SST preference scheme.

We first prove a preliminary result. Let $W_i(X)$ denote the probability that under M player x_i wins the tournament participated in by the set of players X (an SST preference scheme defined on X is assumed).

LEMMA. *Suppose that x_i is stronger than x_j. Then $W_i(X - \{x_j\}) \geq W_j(X - \{x_i\})$. Furthermore, for any $k \neq i, j$*

$$\frac{W_k(X - \{x_j\})}{W_k(X - \{x_i\})} \geq \frac{p_{ki}}{p_{kj}}.$$

Proof. A straightforward induction proof (on the number of rounds) works.

THEOREM. $W_i(X) \geq W_{i+1}(X)$.

Proof. Let T denote the set of all possible realizations of the knockout tournament under the seeding M, and let T_i denote the subset of T in which player x_i wins the tournament. Then

$$W_i(X) = \sum_{t \in T_i} \text{Prob}(t).$$

Let x_l denote the first-round opponent of x_i. Then x_{l-1} is necessarily the first-round opponent of x_{i+1}. In the case that $l = i + 1$, let z denote the winner of the match between x_i and x_{i+1}. Since the relative ranks of z in future rounds remain the same regardless of which player z is, there exists an obvious one-to-one mapping between T_i and T_{i+1} such that if $t \in T_i$ is mapped to $t' \in T_{i+1}$, then t' can be obtained from t by replacing x_i by x_{i+1}. Furthermore, $\mathrm{Prob}(t) \geqslant \mathrm{Prob}(t')$ since $p_{ik} \geqslant p_{i+1,k}$ for all k. It follows that

$$W_i(X) = \sum_{t \in T_i} \mathrm{Prob}(t) \geqslant \sum_{t' \in T_{i+1}} \mathrm{Prob}(t') = W_{i+1}(X).$$

Therefore, we assume $l \neq i + 1$ from now on. We prove the theorem by induction on r, the number of rounds. The theorem is trivially true for $r = 1$. For general r, let F denote any set of first-round winners not counting the matches involving x_i or x_{i+1}. Let $\mathrm{Prob}(F)$ denote the probability that players in F win their first-round matches. Then

$$\begin{aligned}
W_i(X) - W_{i+1}(X) \\
&= \sum_F \big[p_{il} p_{i+1,l-1} W_i(F \cup \{x_i\} \cup \{x_{i+1}\}) \\
&\quad + p_{il} p_{l-1,i+1} W_i(F \cup \{x_i\} \cup \{x_{l-1}\}) \\
&\quad - p_{il} p_{i+1,l-1} W_{i+1}(F \cup \{x_i\} \cup \{x_{i+1}\}) \\
&\quad - p_{li} p_{i+1,l-1} W_{i+1}(F \cup \{x_l\} \cup \{x_{i+1}\}) \big] \mathrm{Prob}(F) \\
&\geqslant \sum_F \big[p_{il} p_{l-1,i+1} W_i(F \cup \{x_i\} \cup \{x_{l-1}\}) \\
&\quad - p_{li} p_{i+1,l-1} W_{i+1}(F \cup \{x_l\} \cup \{x_{i+1}\}) \big] \mathrm{Prob}(F) \quad \text{(by induction)} \\
&\geqslant \sum_F \big[p_{il} p_{l-1,i+1} W_{i+1}(F \cup \{x_{i+1}\} \cup \{x_{l-1}\}) \\
&\quad - p_{li} p_{i+1,l-1} W_{i+1}(F \cup \{x_l\} \cup \{x_{i+1}\}) \big] \mathrm{Prob}(F) \quad \text{(by the Lemma)} \\
&\geqslant \sum_F \left[p_{il} p_{l-1,i+1} \frac{p_{i+1,l-1}}{p_{i+1,l}} - p_{li} p_{i+1,l-1} \right] \\
&\quad \cdot W_{i+1}(F \cup \{x_l\} \cup \{x_{i+1}\}) \mathrm{Prob}(F) \quad \text{(by the Lemma)} \\
&\geqslant 0 \quad (\text{since } p_{il} \geqslant p_{i+1,l} \text{ and } p_{l-1,i+1} \geqslant p_{li}).
\end{aligned}$$

The proof is complete. □

4. Some practical considerations. In practice, one can often infer the SST property of the preference scheme and determine the rank-order of the players. But obtaining an accurate estimate of the probabilities in the preference scheme is a different matter. Fortunately, one need only know the rank-order to apply the method M (also true for S). Furthermore, the "monotone" property of M is assured as long as P is SST. The details of P are never needed.

When the players are human beings who may believe their own versions of rank-orders, different from the official version, we suggest two possible ways of modification. The first is to let the players select their own opponents at each round but to use the official rank-order to determine which player selects first, which second and so on. Of course, players who have been selected by others as opponents lose their right of selection. Presumably, each player will select the weakest player (according to his own version) not yet selected to maximize his probability of winning. In the case that each player agrees with the official version of the rank-order, then the actual pairing would be identical to the one given by M.

A second modification goes one step further by using the official rank-order only in determin-

ing the order of selections for the first round. The orders of selections for future rounds will be completely determined by the outcomes of matches in the previous round. More specifically, the rank of a player in round k is taken to be either his rank or the rank of his opponent in round $k-1$, whichever is higher. This modification will be suitable if the official version is put together with little merit or confidence, and the general guideline is to let the results speak for themselves.

References

1. F. R. K. Chung and F. K. Hwang, Do stronger players win more knockout tournaments? J. Amer. Statist. Assoc., 73 (1978) 593–596.
2. H. A. David, The Methods of Paired Comparisons, Hafner, New York, 1963.
3. J. D. Gibbons, I. Olkin and M. Sobel, Selecting and Ordering Populations: A New Statistical Methodology, Wiley, New York, 1977.

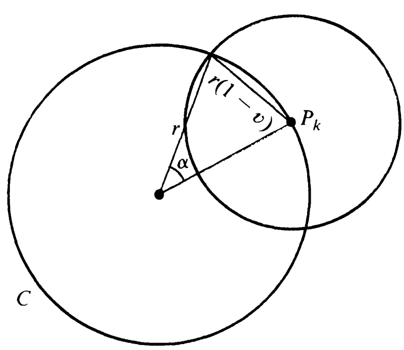
ON BRIDGES' PROOF OF THE GREAT PICARD THEOREM

ANDREANA S. MADGUEROVA
Institute of Mathematics, P. O. Box 373, Sofia -1090, Bulgaria

The proof of the Lemma in [**1**] can be simplified considerably by means of the following result.

If $0 < v \leqslant 0.12$, *if* C *is any circle, of radius* r, *in the complex plane, and if* P *is any point of* C, *then there exist six points* $P_1 = P,\ldots,P_6$ *of* C *such that the discs* $B(P_k, r(1-v))$ *cover* C, *and* $P_{k+1} \in B(P_k, r(1-v))$ *for* $k = 1,\ldots,5$.

Proof. This follows from an elementary trigonometric argument, using the fact that, if $0 < v \leqslant 0.12$, $\sin(\alpha/2) = (1-v)/2$, then $7\alpha > 2\pi$, (see Fig. 1).



$\blacksquare$

FIG. 1

Now let $0 < v \leqslant 0.12$. According to Schottky's Theorem, to each $\alpha > 0$ there corresponds $C(\alpha) > 0$ such that, if f is analytic and omits the values 0, 1 in a region containing $\overline{B}(a,(1-v)/2)$, and if $|f(a)| \leqslant \alpha$, then $|f(z)| \leqslant C(\alpha)$ whenever $|z-a| \leqslant (1-v)/2$. The number $\delta(\varepsilon)$ of Bridges' Lemma is now obtained by setting $\delta_1 = C(\varepsilon)$, $\delta_{k+1} = C(\delta_k)$ $(k = 1,\ldots,5)$ and $\delta(\varepsilon) = \delta_6$.

Reference

1. D. S. Bridges, On Montel's proof of the Great Picard Theorem, this MONTHLY, 88 (January 1981) 45–47.

METRIZATION

What purely topological conditions characterize metric spaces? Answers were discovered independently by R. H. Bing, J. Nagata, and Yu. M. Smirnov in 1950 and 1951. The conditions are that the space be regular and that it have a basis that is the countable union of collections of sets, each collection being locally finite (Nagata, Smirnov) or locally discrete (Bing).

The historic photograph above was taken in 1979, in Moscow, by Douglas Cameron; reading from left to right it shows Smirnov, Bing, Nagata. That was the first time the three big names associated with metrization theory had ever been together.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

NONUNIQUENESS AND GROWTH IN FIRST-ORDER DIFFERENTIAL EQUATIONS

WALTER RUDIN
Department of Mathematics, University of Wisconsin–Madison, Madison, WI 53706

The presence or absence of uniqueness in ordinary differential equations of the familiar form

$$y' = f(x, y) \qquad (*)$$

is shown in this note to have some relevance to the qualitative behavior of global solutions at infinity. More specifically, uniqueness implies that the global solutions form a dominated family (Theorem 1), in a sense that will be defined presently, whereas domination can fail quite spectacularly when there is no uniqueness (Theorem 2).

We let R, R^+, R^2 denote the real line, the nonnegative reals, and the plane, respectively. When f: $R^2 \to R$ is continuous, Peano's existence theorem asserts that $(*)$ has at least one solution through every point $(a, b) \in R^2$. In other words, to every (a, b) corresponds at least one function y, defined on some open segment J with $a \in J$, such that $y(a) = b$ and

$$y'(x) = f(x, y(x)) \qquad (x \in J).$$

Equation $(*)$ is said to *have uniqueness* if exactly one solution passes through every point of R^2. This happens, for instance, when f satisfies a Lipschitz condition of order 1 in the second variable, but also in some other situations (see (E) below).

We say that a solution of $(*)$ is *global* if its domain includes R^+. (It is purely a matter of convenience to have R^+ in this definition, rather than R, and nothing essential is lost by this restriction.)

A collection Γ of continuous functions, each taking R^+ into R, will be called *ultimately dominated* if there is a continuous h: $R^+ \to R^+$ with the following property: To every $g \in \Gamma$ corresponds an $x_g \in R^+$ such that $h(x) > |g(x)|$ for all $x > x_g$.

Here are some examples to illustrate these concepts.

(A) The solutions of $y' = y$ are $y = ce^x$. All are global. They are ultimately dominated by $h(x) = xe^x$.

(B) $y' = 1 + y^2$ has no global solution.

(C) The global solutions of $y' = 1 - \exp(-y)$ are

$$y = \log(1 + ce^x) \qquad (c \geqslant 0).$$

(The ones with $c < 0$ are not global.) They are ultimately dominated by $x \log(1 + x)$.

(D) The equation $y' = |y|^{1/2}$ lacks uniqueness. All solutions are global. They are ultimately dominated by x^2.

(E) If f: $R \to R^+$ is continuous and *positive* then the equation $y' = f(y)$ has uniqueness.

To see this, fix $(a, b) \in R^2$, and suppose that $y(a) = b$ and $y'(x) = f(y(x))$ for all x in some segment J, $a \in J$. Let F be the antiderivative of $1/f$ that has $F(b) = a$. By the chain rule,

$$\frac{d}{dx} F(y(x)) = F'(y(x))y'(x) = \frac{y'(x)}{f(y(x))} = 1.$$

Thus $F(y(x)) - x$ is constant on J. Since $F(b) = a$ and $y(a) = b$, this constant is 0. Since F is strictly increasing, we conclude that $y(x) = F^{-1}(x)$ on J.

Theorem 1 is a simple consequence of the following:

LEMMA. *Every countable collection of continuous functions taking R^+ into R is ultimately dominated.*

Proof. Assume g_n: $R^+ \to R$ is continuous, for $n = 1,2,3,\ldots$. Define $\psi(x) = 0$ if $x < 0$, $\psi(x) = x$ if $0 \leq x \leq 1$, $\psi(x) = 1$ if $x > 1$, and put

$$h(x) = 1 + \sum_{n=1}^{\infty} \psi(x-n)\,|\,g_n(x)\,|\,.$$

On any bounded subset of R^+, only finitely many summands are $\neq 0$. Hence h is continuous. If $x > n+1$, then $h(x) > |\,g_n(x)\,|$.

THEOREM 1. *If equation $(*)$ has uniqueness, then the set Γ of its global solutions is ultimately dominated.*

Proof. Assume Γ nonempty, without loss of generality. Let E be the set of all $c \in R$ such that $y(0) = c$ for some $y \in \Gamma$. There is an at most countable set $E_0 \subset E$ such that every member of $E \setminus E_0$ lies between two members of E_0. Let Γ_0 be the set of all $y \in \Gamma$ with $y(0) \in E_0$.

Since $(*)$ has uniqueness, no two solution curves intersect. Every $y \in \Gamma \setminus \Gamma_0$ lies, therefore, between two members of Γ_0. The lemma, applied to Γ_0, shows, therefore, that Γ is ultimately dominated.

THEOREM 2. *There is a continuous f: $R \to R^+$ with the following properties*:

(I) *Every solution of $y' = f(y)$ is in C^∞.*

(II) *To every continuous h: $R^+ \to R^+$ corresponds a global solution y of $y' = f(y)$ such that $y(x) > h(x)$ for all $x \in R^+$.*

We recall that C^∞ is the class of all infinitely differentiable functions; similarly, for $p = 1,2,3,\ldots,$ C^p is the class of all functions whose pth derivative is continuous.

Note that (II) implies (as a rather weak corollary) that the collection of all global solutions of $y' = f(y)$ is *not* ultimately dominated.

Proof. To make the proof as explicit as possible, we first show that the function f given by

$$f(t) = (1+t^2)\,|\sin \pi t\,|^{1/2} \tag{1}$$

satisfies (II), although (I) holds for it only with C^1 in place of C^∞.

The significant properties of f are

(i) $f(n) = 0$ for all integers n,

(ii) $f(t) > 0$ for all other $t \in R$,

(iii) $\int_{-\infty}^{\infty}[1/f(t)]\,dt < \infty$.

Define

$$F(t) = \int_{-\infty}^{t} \frac{ds}{f(s)} \qquad (t \in R). \tag{2}$$

This F maps R onto $(0,\delta)$ for some $\delta < \infty$, F is strictly increasing, and $F'(n) = +\infty$ for all integers n. Let $\Phi = F^{-1}$, the inverse of F. Then

(α) Φ maps $(0,\delta)$ onto R,

(β) Φ is strictly increasing,

(γ) $\Phi'(x) = 0$ for those (and only those) values of x for which $\Phi(x)$ is an integer.

Also, $x = F(\Phi(x))$ gives $1 = F'(\Phi(x))\Phi'(x)$, and, since $F' = 1/f$, it follows that

$$\Phi'(x) = f(\Phi(x)) \qquad (0 < x < \delta). \tag{3}$$

Hence Φ is a (nonglobal) solution of $y' = f(y)$. The same is true of its translates $\Phi(x - c)$, for arbitrary $c \in R$.

Now consider a point $(a, b) \in R^2$. Choose c so that $\Phi(a - c) = b$, and put $\phi(x) = \Phi(x - c)$. If b is not an integer, then, locally, ϕ is the only solution of $y' = f(y)$ through (a, b); this follows from a local version of (E), or from the fact that f satisfies a Lipschitz condition of order 1 in a neighborhood of b. But if b is an integer, there is another solution, namely the constant b, since $f(b) = 0$. Moreover, we can choose either ϕ or b just to the left of a, and either ϕ or b just to the right of a, thus getting four possibilities at (a, b). This amount of flexibility makes it easy to prove (II):

Let $h: R^+ \to R^+$ be continuous. Assume h to be strictly increasing to ∞, without loss of generality.

Choose t_i, for $i = 1, 2, 3, \ldots$, so that $t_1 > \delta$, $t_{i+1} > \delta + t_i$, and so that $h(\delta + t_i)$ is a positive integer, say n_i. Put $s_0 = 0$, and then choose $c_1, s_1, c_2, s_2, \ldots$, so that

$$\Phi(t_i - c_i) = n_i, \quad \Phi(s_i - c_i) = n_{i+1}. \tag{4}$$

Obviously, $t_i < s_i$. By (α), $s_i < \delta + t_i$, so that $s_i < t_{i+1}$. We can therefore define a global solution of $y' = f(y)$ by setting

$$y(x) = \begin{cases} n_i & (s_{i-1} \leqslant x \leqslant t_i) \\ \Phi(x - c_i) & (t_i \leqslant x \leqslant s_i) \end{cases} \tag{5}$$

for $i = 1, 2, 3, \ldots$. If $s_{i-1} \leqslant x \leqslant s_i$, then

$$h(x) \leqslant h(s_i) < h(\delta + t_i) = n_i \leqslant y(x). \tag{6}$$

Thus $y(x) > h(x)$ for all $x \in R^+$, and (II) is proved.

That (I) holds, with C^1 in place of C^∞, is obvious. Higher differentiability of the solutions can be assured as follows. Let p be a positive integer, and replace the exponent $1/2$ in (1) by $1 - 1/p$. It is then easy to check that $\Phi \in C^{p-1}$ and that, in fact, $\Phi' = \Phi'' = \cdots = \Phi^{(p-1)} = 0$ wherever Φ assumes integer values. Everything else stays the same.

To obtain C^∞ solutions, one can replace (1) by

$$f(t) = (1 + t^2)\,|\sin \pi t|\log^2|\sin \pi t|, \tag{7}$$

or (to avoid the verification that the solutions really are in C^∞) one can turn the construction around and start by building a strictly increasing C^∞-function Φ that takes $(0, 1)$ onto R, such that *all* derivatives of Φ are 0 wherever Φ assumes integer values. There is then a unique f that satisfies

$$f(\Phi(x)) = \Phi'(x) \qquad (x \in R); \tag{8}$$

obviously, Φ is a solution of $y' = f(y)$. The rest of the proof is as above.

To conclude, and to provide some background, we briefly describe what is known along the lines of Theorem 2 for *algebraic* differential equations, i.e., for equations of the form

$$P(x, y, y', \ldots, y^{(n)}) = 0 \tag{$**$}$$

where P is a polynomial in $n + 2$ variables, with real coefficients, and n is the order of the equation.

When $n = 1$, Borel [**3**, p. 27] proved that all global solutions of $(**)$ are ultimately dominated by $\exp(e^x)$; more precise bounds were obtained by Lindelöf [**5**] and Hardy [**4**]. No such universal bounds exist when $n \geqslant 2$; see [**2**], where, corresponding to every *preassigned* continuous $h: R^+ \to R^+$, a function y is constructed that satisfies an equation $(**)$ with $n = 2$, such that $y(x_i) > h(x_i)$ for some *sequence* $\{x_i\}$ that tends to $+\infty$. Bank [**1**] obtained a monotonic y with these properties, but only for $n \geqslant 3$. Rubel [**6**] has recently shown that there is an equation $(**)$, with $n = 4$ (and x absent), which has the following universal approximation property: If ϕ and ε are continuous

functions on R, $\varepsilon > 0$, then $(**)$ has a solution y such that $|\phi(t) - y(t)| < \varepsilon(t)$ for all $t \in R$.

Borel's theorem shows that the phenomenon described by Theorem 2 exists for no algebraic differential equation of order 1. Whether it exists for algebraic equations of order 2 or 3 seems to be unknown.

Note. The referee has pointed out that there is an earlier, very complicated example (starting on p. 18 of [7]; see also [8]) of an equation $y' = f(x, y)$, with f continuous, for which uniqueness fails at *every* point of R^2. This equation does not, however, seem to have the property asserted by Theorem 2 (II).

This research was partially supported by NSF Grant MCS 78-06860, and by the William F. Vilas Trust Estate.

References

1. S. Bank, Some results on analytic and meromorphic solutions of algebraic differential equations, Adv. in Math., 15 (1975) 41–62.

2. N. Basu, S. Bose, and T. Vijayaraghavan, A simple example for a theorem of Vijayaraghavan, J. London Math. Soc., 12 (1937) 250–252.

3. E. Borel, Mémoire sur les séries divergentes, Ann. Sci. École Norm. Sup., 16 (1899) 9–136.

4. G. H. Hardy, Some results concerning the behavior at infinity of a real and continuous solution of an algebraic differential equation of the first order, Proc. London Math. Soc., 10 (1912) 451–468.

5. E. Lindelöf, Sur la croissance des intégrales des équations différentielles algébriques du premier ordre, Bull. Soc. Math. France, 27 (1899) 205–215.

6. L. A. Rubel, A universal differential equation, preprint.

7. P. Hartman, Ordinary Differential Equations, Wiley, New York, 1964.

8. M. Lavrentieff, Sur une équation différentielle du premier ordre, Math. Zeit., 23 (1925) 197–209.

LINEAR DIFFERENTIAL EQUATIONS ON THE COMPLEX PLANE

A. K. BOSE

Department of Mathematical Sciences, Clemson University, Clemson, SC 29631

Introduction. Consider a homogeneous linear differential equation of order n

$$y^{(n)} = a_{n-1}y^{(n-1)} + a_{n-2}y^{(n-2)} + \cdots + a_0 y, \tag{1}$$

where a_k, $k = 0, 1, 2, \ldots, n-1$, are functions continuous in a region (open connected set) R of the complex plane.

Let S_R^n be the solution set of (1) over the region R; S_R^n is a linear space of analytic functions over the field of complex numbers.

The purpose of this note is to emphasize the fact that, unlike homogeneous linear differential equations on the real line, the continuity of the coefficient functions a_k, $k = 0, 1, 2, \ldots, n-1$, does not imply that dimension of S_R^n (dim S_R^n) is n. In fact,

$$\dim(S_R^n) < n$$

if at least one of the coefficients a_k is not analytic in R. This fact can be established by proving the following theorem:

THEOREM 1. *For an arbitrary region R, where all the coefficient functions a_k of equation* (1) *are continuous*, $\dim(S_R^n) = n$ *if and only if each a_k, $k = 0, 1, \ldots, n-1$, is analytic in R.*

To prove Theorem 1, we need the following lemmas about S_R^n:

LEMMA 1. *If f is a solution of* (1) *having a zero of order n, then $f \equiv 0$ for all z in R.*

The usual proof of Lemma 1 in the case of homogeneous linear differential equations on the real line is not applicable here, so that we need to give a different proof.

From the hypothesis of Lemma 1 and from (1), it is clear that f has a zero, say at $z_0 (z_0 \in R)$,

of order at least $n+1$. If $f \not\equiv 0$, then there is a positive integer $p \geqslant 1$ such that f has a zero of order $n+p$ at z_0. Hence, in some neighborhood D of z_0

$$f(z)=(z-z_0)^{n+p}g(z),$$

where $g(z)$ is analytic in D and $g(z_0) \neq 0$.

Taking successive derivatives of $f(z)=(z-z_0)^{n+p}g(z)$ and substituting in (1), we get an equation of the form

$$(n+p)(n+p-1)\cdots(p+1)(z-z_0)^p g(z)=(z-z_0)^{p+1}k(z),$$

where $k(z)$ is continuous in D. Therefore,

$$(n+p)(n+p-1)\cdots(p+1)g(z)=(z-z_0)k(z)$$

is true in some deleted neighborhood of z_0. Taking the limit as $z \to z_0$ we get

$$(n+p)(n+p-1)\cdots(p+1)g(z_0)=0,$$

so that $g(z_0)=0$, implying a contradiction. Hence $f \equiv 0$.

LEMMA 2.

$$\dim(S_R^n) \leqslant n.$$

Suppose that $\dim(S_R^n)>n$, and let $y_1, y_2, \ldots, y_{n+1}$ be $n+1$ linearly independent members of S_R^n. Since the system of n linear equations with $n+1$ unknowns and $z_0 \in R$,

$$\sum_{k=1}^{n+1} x_k y_k^{(i)}(z_0)=0, \quad i=0,1,2,\ldots,n-1,$$

has a nontrivial solution $(c_1, c_2, \ldots, c_{n+1})$, it follows that the solution $c_1y_1+c_2y_2+\cdots+c_{n+1}y_{n+1}$ of (1) has a zero of order n at z_0. Hence, by Lemma 1,

$$c_1y_1+c_2y_2+\cdots+c_{n+1}y_{n+1} \equiv 0 \text{ in } R,$$

contradicting the fact that $y_1, y_2, \ldots, y_{n+1}$ are linearly independent in R.

Proof of Theorem 1. We shall skip the proof of the "if part" of Theorem 1, as it is well known that if all the coefficient functions a_k, $k=0,1,2,\ldots,n-1$, are analytic in R, then $\dim(S_R^n)=n$. So, here we shall give the proof of the "only if" part. That is, $\dim(S_R^n)=n$ implies that all the coefficient functions a_k, $k=0,1,2,\ldots,n-1$, are analytic in R. This we shall prove by applying mathematical induction on the order n of equation (1). Consider a first-order homogeneous equation

$$y'=a(z)y, \tag{2}$$

where $a(z)$ is a function continuous in the region R. Let S_R^1 be the solution set of (2). Suppose that $\dim(S_R^1)=1$ and that $f(z)$ is a nontrivial solution of (2). Consider a $z_0 \in R$ for which $f(z_0) \neq 0$. Then $f'(z)/f(z)=a(z)$ is analytic in some neighborhood of z_0. That is, $a(z)$ is analytic at each point z in R where $f(z) \neq 0$. Since the zeros of a nontrivial analytic function are isolated, and since $a(z)$ is continuous in R, it follows that $a(z)$ is analytic in R. Hence the theorem is true for $n=1$.

Assume now that the theorem is true for some positive integer n. Consider a homogeneous linear differential equation of order $n+1$,

$$y^{(n+1)}=b_ny^{(n)}+b_{n-1}y^{(n-1)}+\cdots+b_0y, \tag{3}$$

where each b_k, $k=0,1,2,\ldots,n$, is a function continuous in R. Let S_R^{n+1} be the solution set of (3), and let $\dim(S_R^{n+1})=n+1$.

Let $y_1, y_2, \ldots, y_{n+1}$ be a basis of S_R^{n+1}. Let $z_0 \in R$ be such that $y_1(z_0) \neq 0$. Let D be a neighborhood of z_0 such that $y_1(z) \neq 0$ in D. Then each of the n functions y_2/y_1,

$y_3/y_1,\ldots,y_{n+1}/y_1$ is analytic in D. Let $Y_2, Y_3,\ldots, Y_{n+1}$ be n functions analytic in D such that

$$Y_k = (y_k/y_1)', \quad k = 2,3,\ldots,n+1.$$

It is easy to see that $Y_2, Y_3,\ldots, Y_{n+1}$ are also linearly independent in D. Now by reducing the order of (3) by one, we see that each Y_k, $k = 2,3,\ldots,n+1$, is a solution of the nth order homogeneous equation

$$u^{(n)} = c_{n-1}u^{(n-1)} + c_{n-2}u^{(n-2)} + \cdots + c_0 u \tag{4}$$

in D with coefficients $c_{n-1}, c_{n-2},\ldots,c_0$, continuous in D, where

$$c_{n-1} = y_1^{-1}\left[-\binom{n+1}{1}y_1' + b_n y_1\right]$$

$$c_{n-2} = y_1^{-1}\left[-\binom{n+1}{2}y_1'' + \binom{n}{1}b_n y_1' + b_{n-1}y_1\right]$$

$$c_{n-3} = y_1^{-1}\left[-\binom{n+1}{3}y_1''' + \binom{n}{2}b_n y_1'' + \binom{n-1}{1}b_{n-1}y_1' + b_{n-2}y_1\right]$$

$$\cdots$$

$$c_0 = y_1^{-1}\left[-\binom{n+1}{n}y_1^{(n)} + \binom{n}{n-1}b_n y_1^{(n-1)} + \binom{n-1}{n-2}b_{n-1}y_1^{(n-2)} + \cdots + b_1 y_1\right].$$

Let S_D^n be the solution set of (4) in D. Since $Y_2, Y_3,\ldots, Y_{n+1}$ are linearly independent in D, it follows that $\dim(S_D^n) = n$. Therefore, by the induction hypothesis each of $b_n, b_{n-1},\ldots,b_1$ is analytic in D. Again

$$y_1^{(n+1)} = b_n y_1^{(n)} + \cdots + b_0 y_1 \quad \text{for all } z \in D$$

implies that b_0 is analytic in D. Hence each of the coefficients b_k, $k = 0,1,2,\ldots,n$, is analytic at each point z in R such that $y_1(z) \neq 0$. Since zeros of y_1 are isolated and each b_k is continuous in R, it follows that each b_k is analytic in R. This completes the proof.

Examples. On the complex plane C consider the solution sets of the following equations:

(1) $y' = |z|y, \dim(S_C^1) = 0$.

(2) $y'' = z|z|y' - |z|y, \dim(S_C^2) = 1$, where $y = z$ is a solution of (2).

(3) $y'' = |z|y' + |z|y, \dim(S_C^2) = 0$.

(4) $y''' = |z|zy'' - |z|y', \dim(S_C^3) = 2$.

We see that $y = 1$ and $y = z^2/2$ are two linearly independent solutions of (4).

References

1. E. A. Coddington, An Introduction to Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, N. J., 1961.

2. E. T. Copson, An Introduction to the Theory of Functions of a Complex Variable, Oxford University, Oxford, 1935.

3. E. L. Ince, Ordinary Differential Equations, Longmans, Green, London, 1927; Dover, New York, 1956.

EVALUATION OF CERTAIN REAL INTEGRALS BY CONTOUR INTEGRATION

BALDEV K. SACHDEVA AND BERTRAM ROSS
Department of Mathematics, University of New Haven, West Haven, CT 06516

1. The usual (textbook) approach for evaluating integrals of the form $\int_0^\infty x^m\, dx/(x^n + a)$ $(a > 0)$ by the method of residues is needlessly restricted by the following assumptions: (i) m and n are even integers; (ii) the region of integration is the upper half-plane, which requires the

evaluation of residues of the integrand at all its poles in the upper half-plane. It was pointed out by Farrell and Ross [1], and Melamed and Kaufman [2], that the restriction (i) is unnecessary and that the half-plane of integration can be replaced by a sectoral contour enclosing only one pole of the integrand. This simplifies the analysis, for it requires the calculation of one residue rather than the calculation of all the residues in the upper half-plane.

The present note extends the idea of a sectoral contour to the evaluation of real integrals of the form $\int_0^\infty x^p f(x)\,dx$ where $f(x)$ is a rational function of x^q (q, a positive integer), having poles of order higher than one. It suffices to consider integrals of the form:

$$I=\int_0^\infty \frac{x^p\,dx}{[P(x)]^n} \tag{1}$$

where $P(x)$ is a polynomial of degree m in x^q and has simple zeros; q, n are positive integers, and $0<p+1<mnq$. Multiple zeros lead to poles of order higher than n and can be handled with minor modification by the method of Section 2.

We introduce a sectoral contour Γ consisting of (i) the segment of the real axis from $x=0$ to $x=R$; (ii) the arc A_R of the circle $|z|=R$ to the ray $\arg z=2\pi/q$; (iii) the ray $z=r\exp(2\pi i/q)$ from $r=R$ to the origin. We then have

$$\int_\Gamma \frac{z^p\,dz}{[P(z)]^n}=2\pi i\sigma \qquad (p>0) \tag{2}$$

where σ is the sum of the residues at the poles of the integrand lying inside Γ. Writing the integral along Γ and noting that in the limit $R\to\infty$ the integral

$$\int_{A_R} \frac{z^p\,dz}{[P(z)]^n}\to 0,$$

we obtain

$$[1-\exp\{2\pi i(p+1)/q\}]\int_0^\infty \frac{x^p\,dx}{[P(x)]}=2\pi i\sigma. \tag{3}$$

When $-1<p<0$, the contour Γ is indented at the origin; equation (3) is still valid because the integral along the indentation vanishes in the limit.

2. Calculation of Residues. Consider the meromorphic function

$$f(z)=\frac{Q(z)}{[P(z)]^n}$$

where $P(z)$ is a polynomial of degree m. Let $z=\lambda$ be a pole of order n for the function $f(z)$. To compute the residue of $f(z)$ at $z=\lambda$ we write

$$f(z)=\frac{\sum_{k=0}^{\infty} a_k(z-\lambda)^k}{\left\{(z-\lambda)\sum_{k=0}^{m-1} b_k(z-\lambda)^k\right\}^n}$$

$$=\frac{1}{(z-\lambda)^n}\left\{\sum_{k=0}^{\infty} a_k(z-\lambda)^k\right\}\left\{\sum_{k=0}^{m-1} b_k(z-\lambda)^k\right\}^{-n}, \tag{4}$$

where

$$a_k=\frac{Q^{(k)}(\lambda)}{k!}, \tag{5}$$

$$b_k = \frac{P^{(k+1)}(\lambda)}{(k+1)!}\,;\; b_k = 0 \quad \text{for } k > m-1. \tag{6}$$

We now use the multinomial formula [3], which states that, for any real number s,

$$\left\{\sum_{k=0}^{\infty} b_k(z-z_0)^k\right\}^s = \sum_{k=0}^{\infty} B_k(z-z_0)^k, \tag{7}$$

where the coefficients B_k and b_k are related by

$$B_0 = b_0^s,$$
$$B_k = \frac{1}{kb_0}\sum_{j=1}^{k}[j(s+1)-k]b_jB_{k-j}. \tag{8}$$

Using (7), (8) with $s=-n$ in equation (4), we have

$$f(z) = \frac{1}{(z-\lambda)^n}\left\{\sum_{k=0}^{\infty} a_k(z-\lambda)^k\right\}\left\{\sum_{k=0}^{\infty} B_k(z-\lambda)^k\right\}$$
$$= \frac{1}{(z-\lambda)^n}\sum_{k=0}^{\infty} C_k(z-\lambda)^k \tag{9}$$

where

$$C_k = \sum_{j=0}^{k} A_jB_{k-j}, \tag{10}$$

and

$$B_k = \frac{1}{kb_0}\sum_{j=1}^{k}[j(-n+1)-k]b_jB_{k-j}. \tag{11}$$

From (9), which is the Laurent expansion of $f(z)$,

$$\text{the residue of } f(z)(\text{at } z=\lambda) \text{ is } C_{n-1}. \tag{12}$$

3. By way of illustration we evaluate the integral

$$\int_0^{\infty}\frac{\sqrt{x}\,dx}{(x^4+1)^3}.$$

Here

$$f(z) = \frac{z^{1/2}}{(z^4+1)^3} = \frac{Q(z)}{[P(z)]^3}, \qquad 0 \leqslant \arg z \leqslant \frac{\pi}{2} \tag{13}$$

and the contour Γ is the sector $0 \leqslant \arg z \leqslant \pi/2$, which contains only one pole $z=\lambda=\exp(i\pi/4)$ of order 3, with residue $C_2 = a_0B_2 + a_1B_1 + a_2B_0$. Here

$$\begin{aligned}
&a_0 = \lambda^{1/2}, \quad a_1 = \frac{1}{2}\lambda^{-1/2}, \quad a_2 = -\frac{1}{8}\lambda^{-3/2},\\
&b_0 = 4\lambda^3, \quad b_1 = 6\lambda^2, \quad b_2 = 4\lambda,\\
&B_0 = (b_0)^{-3} = \lambda^{-9}/64,\\
&B_1 = \frac{1}{b_0}(-3)b_1B_0 = -9\lambda^{-10}/128, \quad B_2 = \frac{1}{2b_0}[-4b_1B_1 - 6b_2B_0] = 21\lambda^{-11}/128,
\end{aligned} \tag{14}$$

whence from (14):

$$C_2 = \frac{65}{512}\lambda^{-21/2} = \frac{65}{512}\exp(-21\pi i/8).$$

From (3) we get

$$\int_0^\infty \frac{\sqrt{x}\,dx}{(x^4+1)^3} = \frac{2\pi i}{\{1-\exp(3\pi i/4)\}}\,\frac{65}{512}\cdot\exp(-21\pi i/8)$$
$$= \frac{65}{512}\quad \mathrm{Csc}\left(\frac{3\pi}{8}\right).$$

References

1. Orin J. Farrell and Bertram Ross, Note on evaluating certain real integrals by Cauchy's residue theorem, this MONTHLY, 68 (1961) 151–152.
2. S. Melamed and H. Kaufman, Evaluation of certain improper integrals by residues, this MONTHLY, 72 (1965) 1111–1112.
3. H. W. Gould, Coefficient identities for powers of Taylor and Dirichlet series, this MONTHLY, 81 (1974) 3–14.

A FUNCTION WITH A DISCONTINUOUS DERIVATIVE

BRUCE PETERSON
Department of Mathematics, Middlebury College, Middlebury, VT 05753

The function

$$f(x) = \begin{cases} x^2\sin\dfrac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is the standard example of a function with a discontinuous derivative [**1**, p. 36]. Variations on the theme, usually involving functions of the form $\sin r(x)$ where $r(x)$ has an asymptote at the origin, provide several other examples of extreme behavior on the part of a derivative.

In all these examples $f'(0) = 0$, but the range of the derivative in each interval $(0,\delta)$ is an interval with 0 as an interior point. In the case of $f(x)$ the derivative takes on all the values in $[-1,1]$ in any interval $(0,\delta)$. This observation may lead the student to ask whether it is possible to find a function $k(x)$ with discontinuous derivative such that, say,

$$k'(x)\begin{cases} >0 & \text{if } x \neq 0 \\ =0 & \text{if } x = 0. \end{cases}$$

The intermediate-value property does not rule this out, but examples based on the $\sin(1/x)$ model fail because such functions have a 0-derivative infinitely often in every interval $(0,\delta)$.

The desired functions do exist however. By piecing together a number of simple curves lying close to the graph of $y = x^2$, we can find a function with the desired properties. The procedure is as follows: First define, for $n = 2,3,4,\ldots$:

$$h_n(x) = \left\{\begin{array}{ll} -K_n\left(-x+\dfrac{3}{2^{n+1}}\right)^{2^n/3} + \dfrac{5}{2^{2n+1}} & \text{if } \dfrac{1}{2^n} \leqslant x \leqslant \dfrac{3}{2^{n+1}} \\ K_n\left(x-\dfrac{3}{2^{n+1}}\right)^{2^n/3} + \dfrac{5}{2^{2n+1}} & \text{if } \dfrac{3}{2^{n+1}} \leqslant x \leqslant \dfrac{1}{2^{n-1}} \end{array}\right|$$

where $K_n = 3 \cdot 2\,[(2^n/3)(n+1) - 2n - 1]$.

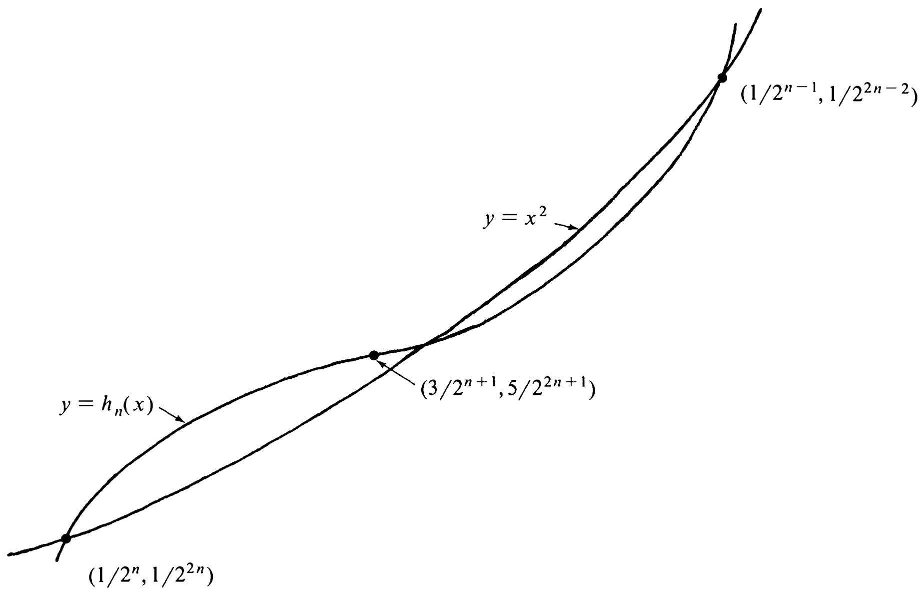


FIG. 1

It is easy to verify that $h_n(x)$ is strictly increasing on $[1/2^n, 1/2^{n-1}]$ and that

$$h_n(1/2^n) = 1/2^{2n},\ h_n(1/2^{n-1}) = 1/2^{2n-2},$$
$$h_n'(1/2^n) = h_n'(1/2^{n-1}) = 1, \quad h_n'(3/2^{n+1}) = 0.$$

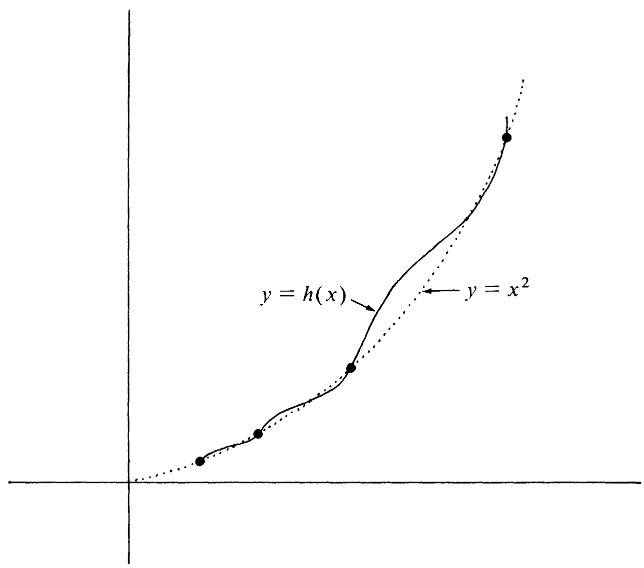


FIG. 2. A portion of $h(x)$, vertical scale exaggerated.

[Continued on p. 263.]

That is, the graph of $h_n(x)$ coincides with that of $y = x^2$ at the end points of the interval $[1/2^n, 1/2^{n-1}]$, has slope 1 at each end point, and has a point of inflection at the midpoint of the interval, where the slope is 0. Now, for $x \geqslant 0$, define

$$h(x) = \begin{cases} 0 & \text{if } x = 0 \\ h_n(x) & \text{if } 1/2^n \leqslant x < 1/2^{n-1}, \quad n = 2, 3, \ldots \\ x - \dfrac{1}{4} & \text{if } x \geqslant \dfrac{1}{2} \end{cases}$$

and for $x < 0$ define $h(x) = -h(-x)$. Restricting our attention to positive x since the curve is symmetric in the origin, we have, for $n = 2, 3, \ldots$,

$$h(1/2^n) = 1/2^{2n}, \quad h'(1/2^n) = 1, \quad h'(3/2^{n+1}) = 0.$$

For the derivative h', $h'(x) \geqslant 0$, and, in fact, h is a strictly increasing function. Since $h'(x)$ takes on all values in $[0, 1]$ in every interval $(0, \delta)$, the derivative $h'(x)$ fails to be continuous at the origin. It is easy to verify that $h'(0) = 0$. Finally, the function

$$k(x) = \begin{cases} -x^2 + h(x) & \text{if } x < 0 \\ x^2 + h(x) & \text{if } x \geqslant 0 \end{cases}$$

has all the desired properties.

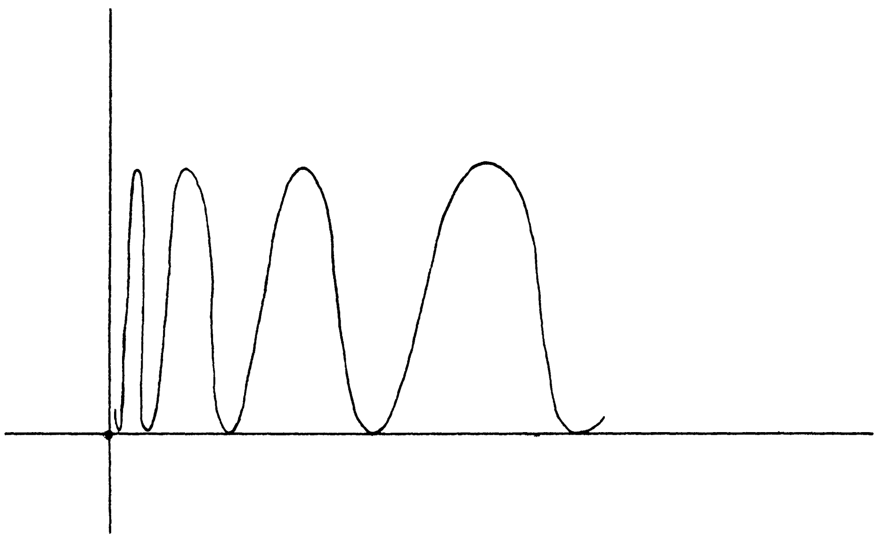


FIG. 3. $h'(x)$ $(h'(0) = 0)$

Reference

1. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.

MISCELLANEA

70. Everybody praises the incomparable power of the mathematical method, but everybody is also aware of its incomparable unpopularity.

—J. Rosanes, *Jber. Deutsch. Math. Verein.*, 13 (1904) 17.

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S(15-17). Science, Technology, and National Policy. Ed: Thomas J. Kuehn, Alan L. Porter. Cornell U Pr, 1981, 530 pp, $9.95 (P); $35. [ISBN: 0-8014-9876-7; 0-8014-1343-5] A reader on technology and public policy, intended as a source book for courses on this topic. 25 selections arranged in two parts: social context of science and technology, and the role of science and technology in American government. Although all of these essays date from the expansive ′60′s and ′70′s, the current climate of retrenchment in public scientific priorities lends urgency to the issues discussed in this volume. LAS

General, P, L. From Being to Becoming: Time and Complexity in the Physical Sciences. Ilya Prigogine. WH Freeman, 1980, xix + 272 pp, $12.95 (P). [ISBN: 0-7167-1108-7] A philosophical study of the implications of the author′s Nobel Prize winning work in chemistry. Intended for a general (scientist) reader with some background in physical chemistry and thermodynamics, this book shows how systems far from equilibrium evolve elaborate structures: order emerges out of chaos. JAS

General, L. Computers For Everybody. Jerry Willis, Merl Miller. Dilithium Pr, 1981, viii + 173 pp, $4.95 (P). [ISBN: 0-918398-49-5] Comprehensive, non-technical guide to purchasing and beginning to use a home computer for neophytes. Explanations of computer components and peripherals, ideas for possible uses, suggestions for establishing purchase criteria. Lists of publications and software distributors and descriptions of most popular systems. Good first book. MW

General, S(13). Barron′s How to Prepare for the College-Level Examination Program (CLEP) General Examinations, Third Edition. Ed: William C. Doster, et al. Barron′s Educ Ser, 1979, 566 pp, $6.50 (P). [ISBN: 0-8120-2011-1] Designed for a variety of purposes, the CLEP General Examinations cover five broad areas, one of which is mathematics. The mathematics examination has two parts, basic skills and concepts, and content (sets, logic, real number system, functions and their graphs, probability and statistics, and miscellaneous topics). This guide has review questions and explanations, and four complete sample examinations for each area. RSK

General, L. The Facts On File Dictionary of Mathematics. Ed: Carol Gibson. Facts On File, 1981, 216 pp, $14.95. [ISBN: 0-87196-512-7] 1200 entries from school and college mathematics, including computing ("disk," "software") and accounting ("hire purchase"). Major emphasis is on algebra, calculus and descriptive geometry. LAS

Precalculus, T(13). College Algebra and Trigonometry. Michael Sentlowitz, Margaret Trivisone. Addison-Wesley, 1981, xii + 622 pp, $14.95. [ISBN: 0-201-06676-9] Thorough coverage of standard topics; clean format without the cartoons now in vogue. AWR

Precalculus, T(13). Trigonometry: A Complete and Concrete Approach. Harold S. Engelsohn. McGraw-Hill, 1981, x + 470 pp, $17.95. [ISBN: 0-07-019419-X] More trigonometry than most people take time to cover. Nice coverage, with realization that availability and use of hand-held calculators calls for introduction of inverse functions earlier than used to be the case. AWR

Precalculus, T(13: 1). Plane Trigonometry. Richard Miller, Patricia Henry. Brooks/Cole, 1981, ix + 304 pp, $15.95. [ISBN: 0-8185-0421-8] Traditional text with formal writing style, emphasizing calculators (interpolations relegated to appendix). Many sections independent. Circular functions first, but plenty of triangle applications. MW

Precalculus, T(13: 1), S. College Algebra and Trigonometry as Socrates Might Have Taught Them, Third Edition. Robert D. Hackworth, Joseph W. Howland. H & H Pub, 1981, 295 pp, $14.95 (P). Only substantial changes in this Third Edition are a more readable format and the addition of a chapter on systems of equations, matrices, and determinants. No motivation or application of concepts is included in its format of semi-programmed instruction. Socrates′ approach would have been more exciting. (First Edition, TR, May 1979.) JJ

Precalculus, T(13: 1), L. College Algebra with Calculators. Marshall D. Hestenes, Richard O. Hill, Jr. Prentice-Hall, 1982, ix + 406 pp, $20.95. [ISBN: 0-13-140806-2] Although the selection of topics is quite traditional, the authors have made a real (and largely successful) effort to exploit the use of the calculator both to reduce some of the drudgery in calculation and also to make accessible a wider range of solvable problems. The text is brief but imaginative and interesting. Exercises seem interesting and well-chosen. Historical notes for each chapter. JS

Education, S(16-18), P. Selected Issues in Mathematics Education. Ed: Mary Montgomery Lindquist. NCTM, 1980, x + 268 pp, $15 (P). [ISBN: 0-8211-1114-0] Addresses specific issues concerning elementary and secondary mathematics educations, including curriculum, content, instructional aids, student and teacher variables, and national assessment. Not a review of research, it examines the present state of mathematics education and proposes guidelines for future decision making. MW

Education, S(16-18), P. The Psychology of Mathematics for Instruction. Lauren B. Resnick, Wendy W. Ford. Lawrence Erlbaum Assoc, 1981, vi + 266 pp, $24.95. [ISBN: 0-89859-029-9] Review of psychological research on mathematics learning and thinking processes with implications for teaching. Intent is to build a framework for an emerging "psychologically based science of mathematics instruction." Organized around three facets of mathematics--as computation, as a structured discipline, and as problem solving and reasoning. Provides background information on psychological approaches and research methodology, so little background in psychology is required. MW

Education, S(16). The Mathematics Laboratory in the Elementary School: What? Why? and How?. Frank Swetz. Intergalactic Pub, 1981, v + 57 pp, $6.95 (P). [ISBN: 0-936918-03-9] Step-by-step guide to establishing a laboratory including suggested floor plans, equipment lists and storage facilities. Annotated bibliography but no suggestions for student books. Some light printing and several pages out of sequence. MW

Education, P. Source Book of Projects, Science Education, Development and Research. NSF, 1980, xii + 225 pp, (P). A description of FY 1980 grants made by the NSF Division of Science Education Development and Research, with references to earlier awards. Keyword and permuted title indexes help identify projects in special areas. LAS

History, P, L. Philosophy of Mathematics and Deductive Structure in Euclid's Elements. Ian Mueller. MIT Pr, 1981, xv + 378 pp, $37.50. [ISBN: 0-262-13163-3] Intensive examination of the Elements. Principal aims are to thoroughly survey and analyze the contents of the Elements using modern logical symbolism, and to provide an understanding of the Greek conception of foundations of mathematics as contrasted with our own (i.e., Hilbert's geometry, Peano arithmetic, Dedekind reals, etc.). Many of the general conclusions seem to be lost in the detailed analysis of individual definitions and propositions (of which there is a complete list in an appendix). May become a standard reference. GHM

Foundations, P, L. Brouwer's Cambridge Lectures on Intuitionism. Ed: D. van Dalen. Cambridge U Pr, 1981, xii + 109 pp, $22.50. [ISBN: 0-521-23441-7] Posthumous publication of L.E.J. Brouwer's carefully written lecture notes from 1946-1951. This is now the only consolidated, published exposition by Brouwer himself of the fundamental notions of Intuitionism, the school of which he was the founder and chief evangelist. Editorial comments or emendations are not always easy to distinguish from Brouwer's text. GHM

Foundations, T(14-16: 2). Set Theory and Logic. Robert R. Stoll. Dover, 1979, xiv + 474 pp, $6.50 (P). [ISBN: 0-486-63829-4] Unabridged, corrected republication of 1963 edition (W.H. Freeman). Focusses on the mathematics directly relevant to development and study of the real number systems: set theory, construction of number systems, axiomatic theories and formal logic. Until last chapter relatively informal (but rigorous) with many appropriate exercises. Good bridge between undergraduate classroom experience and advanced formal study. GHM

Number Theory, P. Séminaire de Théorie des Nombres, Paris 1979-80: Séminaire Delange-Pisot-Poitou. Ed: Marie-José Bertin. Progress in Math., V. 12. Birkhäuser Boston, 1981, ix + 394 pp, $22. [ISBN: 3-7643-3035-X] An outstanding collection of surveys papers (25 in all) most of which deal with algebraic number theory or diophantine geometry. SG

Linear Algebra, T(15-16: 1). Grundzüge der linearen Algebra. Richard Wagner. Teubner Stuttgart, 1981, 260 pp, (P). [ISBN: 3-519-02756-9] A carefully-written text intended chiefly for prospective secondary teachers. Elementary but sophisticated. JD-B

Algebra, P. Lecture Notes in Mathematics-854: Algebraic K-Theory, Evanston 1980. Ed: E.M. Friedlander, M.R. Stein. Springer-Verlag, 1981, 517 pp, $27 (P). [ISBN: 0-387-10698-7] Proceedings of a conference held at Northwestern University on March 24-27, 1980. JAS

Algebra, T(15-17), S*, L**.** Elements of Algebra and Algebraic Computing. John D. Lipson. Addison-Wesley, 1981, xvi + 342 pp, $34.50. [ISBN: 0-201-04115-4] A text for an (applied) algebra course, not readily comparable to other books. Most of the topics likely to be treated in the usual courses are there, but so, for example, are Pascal procedures for computing in principal ideal domains. Groups, rings, and fields are all there, but many standard results appear as problems because the text presents them in the language of universal algebra and category theory. But don't be misled: this is an elementary presentation based on a two-semester course in the Department of Computer Science at the University of Toronto. JAS

Algebra, S(18), P. Representations of Real Reductive Lie Groups. David A. Vogan, Jr. Progress in Math., v. 15. Birkhäuser Boston, 1981, xvii + 754 pp, $35. [ISBN: 3-7643-3037-6] Assuming a rather thorough familiarity with complex Lie groups and algebras, the author surveys some recent developments in the area of infinite dimensional representations of real reductive Lie groups including work of Langlands, Zuckerman, and Kazhdan-Lusztig. Bibliography, index. JS

Algebra, T(18: 1), S, P, L. Lie Groups, Lie Algebras and Representation Theory. Hans Zassenhaus. Pr U Montreal, 1981, 284 pp, $12 (P). [ISBN: 2-7606-0546-9] An outgrowth of a summer lecture series, this is a fairly elementary and largely self-contained introduction to Lie groups and algebras, with an emphasis on the latter and eventual classification of the simple algebras over the complex field. References. No index. JS

Algebra, P. On Villamayor and Zelinsky's Long Exact Sequence. Mitsuhiro Takeuchi. Memoirs No. 249. AMS, 1981, iv + 178 pp, $10 (P). An extensive development of some special cohomology groups which yield a long exact sequence with a homomorphism of this sequence to the Villamayor and Zelinsky sequence. JAS

Algebra, P. Lecture Notes in Mathematics-867: Séminaire d'Algèbre Paul Dubreil et Marie-Paule Malliavin. Ed: M.P. Malliavin. Springer-Verlag, 1981, v + 476 pp, $22.10 (P). [ISBN: 0-387-10841-6] The proceedings of this seminar for the year 1980. JAS

Algebra, T(18: 1), S, P. Rational Constructions of Modules for Simple Lie Algebras. George B. Seligman. Contemporary Math., V. 5. AMS, 1980, xiii + 185 pp, $10.60 (P). [ISBN: 0-8218-5008-3] A thorough and fairly comprehensive treatment to determine the finite-dimensional irreducible representations of a simple Lie algebra over a field of characteristic zero; "it is sought here to avoid splitting extensions in the realization of the representations, obtaining them by processes that are rational over the given ground field." Approach is "to give a rational procedure for constructing modules of given weight." Considerable background in Lie algebras is necessary. References, no index. JS

Algebra, T(16-18: 1), S, P, L. Modern Algebra: A Constructive Approach. Ian Connell. Elsevier North Holland, 1982, xi + 451 pp, $24.95. [ISBN: 0-444-00609-5] Definitely not just another undergraduate text for a first course, this book might best be used after a semester's exposure to one of the more conventional ones. After a rather austere discussion of the integers, beginning with Peano's axioms, groups and rings are treated briefly, followed by a long chapter on linear algebra. Nearly half of the book is then devoted to Chapter 5 on the real, complex, and p-adic numbers with a broad selection of non-standard topics such as valuations, Hensel's Lemma, the Gelfond-Mazur theorem, and Wedderburn's theorem. Exercises, assignments (both non-trivial), indexes. JS

Algebra, S(18), P. The Structure of Locally Compact Abelian Groups. D.L. Armacost. Pure and Appl. Math., V. 68. Dekker, 1981, vii + 154 pp, $23.75. [ISBN: 0-8247-1507-1] Definitely not for the novice. The reader should be well-schooled in both harmonic analysis and infinite abelian groups (for example by reading Hewitt and Ross and L. Fuchs) before beginning here, but the results are rewarding. Emphasizes algebraic and topological aspects (no measure theory is used) including torsion groups, splitting, connectedness, and homology. "Miscellanea" for each chapter includes exercises, helpful commentary, and research problems. Index, bibliography. JS

Finite Mathematics, T(13: 1). Using Finite Mathematics. Joseph Newmark. Harper & Row, 1982, xv + 604 pp, $19.50. [ISBN: 0-06-385752-9] Presents standard topics in finite mathematics in a nontechnical manner. Includes calculator-oriented problems. The final chapter on computers and Basic is weak and should have been integrated (at the least, connected in some manner) to the content in the rest of the text. JJ

Calculus, T(15), S. Advanced Calculus. Charles Dixon. Wiley, 1981, ix + 147 pp, $28. [ISBN: 0-471-27913-7] Misleading title for those expecting topology of the line and/or plane, limits, theorems on continuity, integration, etc. Deals with vector valued functions of one variable, real valued functions of several variables, and transformations from R^n to R^m. AWR

Calculus, S(13-14), L. A Brief Course of Higher Mathematics. V.A. Kudryavtsev, B.P. Demidovich. Trans: Leonid Levant. MIR Pub, 1981, 693 pp, $16. Elementary calculus--Russian style. Contents are those of like-level U.S. textbooks with a more thorough treatment of analytical geometry. Brief chapter on complex numbers, one on determinants, one on probability and one on linear programming. About 25 pages on Fourier series and boundary value problems. Usual exercises, all with answers, with little emphasis on "story" problems. In depth and rigor, about level with most widely-used U.S. calculus textbooks. Effective sketches, without color. JK

Real Analysis, T(16-17: 2). Lehrbuch der Analysis, Teil 2. Harro Heuser. Teubner Stuttgart, 1981, 736 pp, DM 58 (P). [ISBN: 3-519-02222-2] Second volume (First Volume, TR, November 1981) of a modern and well-written text. Contains chapters on Banach spaces and algebras, the Lebesgue integral, Fourier series, topological spaces, fixed-point theorems and the history of analysis, as well as an extensive treatment of multivariable calculus. JD-B

Real Analysis, P. Limits of Indeterminancy in Measure of T-means of Subseries of a Trigonometric Series. D.E. Men'sov. Proc. of Steklov Inst. of Math. AMS, 1980, 56 pp, $20 (P). [ISBN: 0-8218-3043-0] Existence proof of a universal trigonometric series which represents, in a certain sense, all measurable functions. LAS

Real Analysis, T(16-17: 1, 2), P, L. Measure Theory and Integration. G. de Barra. Halsted Pr, 1981, 239 pp, $64.95. [ISBN: 0-470-27232-5] A carefully structured book which develops integration via measures. Over 300 exercises with fairly detailed solutions in an appendix. The approach is quite standard; can serve both as an introductory text and as a general reference. A rewritten extension of Introduction to Measure Theory (TR, March 1975) by the same author. The price is alarming. PH

Differential Equations, P. Nonlinear Differential Equations: Invariance, Stability, and Bifurcation. Ed: Piero de Mottoni, Luigi Salvadori. Academic Pr, 1981, xi + 357 pp, $22. [ISBN: 0-12-508780-2] Proceedings of a conference held at the Villa Madruzzo, Trento, Italy, May 25-30, 1980. JAS

Differential Equations, T(18), P. Pseudodifferential Operators. Michael E. Taylor. Math. Series, V. 34. Princeton U Pr, 1981, xi + 451 pp, $35. [ISBN: 0-691-08282-0] Building on foundations laid by Hörmander, Taylor develops a theory of pseudodifferential operators using a suitable mixture of functional analysis, Fourier integral operators, energy estimates, and fundamental solutions and parametrics. Includes applications to elliptic and hyperbolic equations, existence and uniqueness. Bibliography, index. JS

Differential Equations, P. Asymptotic Analysis of Singular Perturbations. Wiktor Eckhaus. Stud. in Math. and its Appl., V. 9. Elsevier North-Holland, 1979, xi + 287 pp, $48.75. [ISBN: 0-444-85306-5] An exposition of the deductive structure of singular perturbations--a "fascinating mixture of rigorous analysis, heuristic reasoning and induction from experience." The scope of the book is limited to functions without oscillations, those exhibiting boundary layer behavior. LAS

Differential Equations, P*. Nonlinear Partial Differential Equations in Physical Problems. D. Graffi. Res. Notes in Math., V. 42. Pitman Pub, 1980, 105 pp, $16.95 (P). [ISBN: 0-273-08474-7] Lectures, given in October, 1978, at the Bernoulli Session of the International Centre for Mechanical Sciences, Udine, Italy, on electromagnetism, heat transfer problems with nonlinear boundary conditions, fluid mechanics, electromagnetic fields in a nonlinear plasma and nonlinear optics. Author invites readers to indicate whether and how methods discussed can be made rigorous and what rigorous methods are best suited to the investigation of the nonlinear problems being treated. JK

Differential Equations, S(18), P. Lecture Notes in Mathematics-856: Propagation des Singularités des Solutions d'Equations Pseudo-Différentielles à Caractéristiques de Multiplicités Variables. Richard Lascar. Springer-Verlag, 1981, viii + 237 pp, $11.90 (P). [ISBN: 0-387-10702-9]

Differential Equations, P. Stochastic Stability of Differential Equations. R.Z. Has'minskii. Trans: D. Louvish. Mono. and Textbooks on Mech. of Solids and Fluids, V. 7. Sijthoff & Noordhoff, 1980, xvi + 344 pp. [ISBN: 90-286-0100-7] A translation (with revisions and an appendix and bibliography on recent results) of the 1969 Russian original version. Concerns stability, boundedness, and other properties of solutions to differential equations with stochastic coefficients. Intended for mathematicians and physicists. PZ

Differential Equations, T*(15-17), S, L. An Introduction to Nonlinear Oscillations. Ronald E. Mickens. Cambridge U Pr, 1981, xiv + 224 pp, $69. [ISBN: 0-521-22208-9] A well-designed advanced undergraduate text covering approximation techniques for solving the harmonic oscillator equation with an added "small" nonlinear term. Useful features: some examples are worked out by several methods for comparison; nine pertinent appendices on linear differential equations, Fourier series, asymptotic expansions, etc. PZ

Differential Equations, P. Boundary Value Problems for Elliptic Pseudodifferential Equations. G.I. Eskin. Transl. of Math. Mono., v. 52. AMS, 1981, x + 375 pp, $68. [ISBN: 0-8218-4503-9] Pseudodifferential equations include differential and multi-dimensional singular integral equations as well as integral equations of the first kind and integrodifferential equations whose kernels have a weak singularity. This monograph investigates mixed boundary value problems for elliptic equations. AO

Differential Equations, T*(18: 1, 2), S, P*. Analytic Theory of Partial Differential Equations. David L. Colton. Pitman, 1980, xii + 239 pp, $60. [ISBN: 0-273-08462-3] This text has been designed to provide the reader with a broad overview of the analytic theory of partial differential equations. The last two chapters discuss two classical inverse problems: the backwards heat equation and the inverse scattering problem. AO

Differential Equations, P. Boundary Value Problems of Mathematical Physics, X. Ed: O.A. Ladyženskaja. Proc. of Steklov Inst. of Math., V. 147. AMS, 1980, vi + 213 pp, $64 (P). [ISBN: 0-8218-3002-3) "This collection is devoted to the study of nonlinear partial differential equations, group analysis of the Navier-Stokes equations and boundary layer theory, the construction of functional models of perturbation theory and their application in scattering theory, and the solvability of problems in magnetohydrodynamics." JK

Numerical Analysis, P. Proceedings of the 1981 Army Numerical Analysis and Computers Conference. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC), 1981, xv + 630 pp, (P). Papers from a February 1981 conference at the Redstone Arsenal, Alabama. LAS

Numerical Analysis, T(15-16: 1), S, L. Finite Elements: An Introduction, Volume I. Eric B. Becker, Graham F. Carey, J. Tinsley Oden. Prentice-Hall, 1981, xii + 258 pp, $24.95. [ISBN: 0-13-317057-8] A basic introduction intended for engineering students. Presumes some linear algebra and ordinary differential equations. Proceeds from example to analysis to computer code. Exercises. First of a proposed six volume series. RWN

Numerical Analysis, T(14-15: 1), L. Numerical Mathematics and Computing. Ward Cheney, David Kincaid. Brooks/Cole, 1980, xiv + 362 pp, $21.95. [ISBN: 0-8185-0357-2] Presumes calculus and Fortran. Scientific programming methods such as root finding, integration, linear systems, splines, ordinary differential equations, partial differential equations, linear programming. Some discussion of roundoff and truncation errors. Mathematical and programming exercises. RWN

Numerical Analysis, P. Approximation of Functions by Polynomials and Splines. Ed: S.B. Stečkin. Proc. of Steklov Inst. of Math., No. 145. AMS, 1980, vi + 270 pp, $88 (P). [ISBN: 0-8218-3049-X] Nine research articles. RWN

Numerical Analysis, P. Approximation Theory and Applications. Ed: Zvi Ziegler. Academic Pr, 1981, xi + 358 pp, $26. [ISBN: 0-12-780650-4] Proceedings of a seven-week workshop held at the Technion, Haifa, Israel in May and June of 1980. JAS

Numerical Analysis, S(17-18), P. Finite Modelle gewöhnlicher Randwertaufgaben. Erich Bohl. Teubner Stuttgart, 1981, 318 pp, (P). [ISBN: 3-519-02353-9] On the solution of certain boundary-value problems for second-order ordinary differential equations by finite-difference and finite-element methods. JD-B

Functional Analysis, T(18). Introduction to Linear Operator Theory. Vasile I. Istrăţescu. Pure and Appl. Math., V. 65. Dekker, 1981, xi + 579 pp, $32.50. [ISBN: 0-8247-6896-5] After five chapters covering linear functional analysis, Chapter 6 introduces as the basic notion of the book the concept of the numerical range. AWR

Functional Analysis, P. Lecture Notes in Mathematics-852: Geometry and Probability in Banach Spaces. Laurent Schwartz. Springer-Verlag, 1981, x + 101 pp, $9.80 (P). [ISBN: 0-387-10691-X] Notes by Paul Chernoff of lectures given by Schwartz in Berkeley in April and May 1978, summarizing results from the several preceding years at the Séminaire de l'Ecole Polytechnique in Palaiseau, France that linked recent functional analytic, probabilistic and geometric theories of Banach spaces. LAS

Functional Analysis, T(17), P. Fixed Point Theory: An Introduction. Vasile I. Istrăţescu. Math. and Its Appl., V. 7. D Reidel Pub, 1981, xv + 466 pp, $71. [ISBN: 90-277-1244-7] Presents some standard results (the contraction mapping, Brouwer, Schauder theorems) and some recent results. The goal of making results available to readers of diverse backgrounds is accomplished in a highly readable text written from a functional analysis point of view. Watch out for the price though. AWR

Functional Analysis, T(17-18: 1, 2), S. Functional Analysis. Balmohan Vishnu Limaye. Wiley, 1981, xii + 376 pp, $17.95. [ISBN: 0-470-26933-2] An introductory text with minimal prerequisites; designed for application-oriented courses as well as traditional ones; with exercises. JL

Functional Analysis, T(18), P. Applied Functional Analysis, Second Edition. A.V. Balakrishnan. Appl. of Math., V. 3. Springer-Verlag, 1981, xiii + 373 pp, $34. [ISBN: 0-387-90527-8] A corrected, slightly enlarged version of the First Edition (TR, January 1977) aiming toward significant applications in convex programming, control theory, and stochastic optimization. A rather narrow approach to analysis in Hilbert space, concentrating on semi-groups of operators, either compact or Hilbert-Schmidt type. Bibliography, index. JS

Functional Analysis, P*. Shifts and Periodicity for Right Invertible Operators. D. Przeworska-Rolewicz. Res. Notes in Math., No. 43. Pitman Pub, 1980, 191 pp, $18.95 (P). [ISBN: 0-273-08478-X] Shifts for right invertible operators acting in linear spaces. Applications to functional-differential equations. Conditions of solvability are reduced to a problem of the invertibility of some integral operator. JK

Functional Analysis, T(16-18: 1), S, P, L. Basic Operator Theory. Israel Gohberg, Seymour Goldberg. Birkhäuser Boston, 1981, xiii + 285 pp, $14.95. [ISBN: 3-7643-3028-7] Intended as a text for a first course in operator theory, assuming only a background in linear algebra and advanced calculus. Mostly concerned with Hilbert space, it develops sufficient theory for a proof and effective application of the spectral theorem for compact normal operators; later chapters venture into Banach spaces and compact operators. Applications to integral and differential equations, approximation theory. Many exercises, references, and four appendices. Very reasonably priced. JS

Functional Analysis, T(17-18: 1, 2), S, P. Locally Convex Spaces. Hans Jarchow. Teubner Stuttgart, 1981, 548 pp. [ISBN: 3-519-02224-9] A systematic textbook on locally convex spaces starting with the elementary theory of general topological vector spaces and ending with nuclear spaces. An extensive bibliography is included. AO

Functional Analysis, P. Le distribuzioni nella fisica matematica. Vasilij Sergeevic Vladimirov. MIR, 1981, 320 pp. Translation of a 1979 Russian monograph, published in Moscow. Definition and properties of distributions, integral transforms, applications. LAS

Analysis, S(18), P. Topics in Harmonic Analysis on Homogeneous Spaces. Sigurdur Helgason. Progress in Math., V. 13. Birkhäuser Boston, 1981, ix + 142 pp, $12. [ISBN: 3-7643-3051-1] The homogeneous spaces considered are the Euclidean plane, the two-sphere, and the non-Euclidean plane. "Harmonic analysis" on such spaces is the study of eigenspaces and eigenfunctions of the algebra of left-translation-invariant differential operators. Based on lectures at the Canadian Mathematical Society Summer Seminar, 1980. PZ

Analysis, P. Lecture Notes in Mathematics-866: Mécanique Aléatoire. Jean-Michel Bismut. Springer-Verlag, 1981, xvi + 563 pp, $26.90 (P). [ISBN: 0-387-10840-8] Development and use of the necessary differential analysis "to apply the methods of classical mechanics to a class of stochastic optimization problems, which include classical mechanics systems submitted to random perturbations." JAS

Analysis, P. Equacões Integrais. M.L. Krasnov, A.I. Kisseliov, G.I. Makarenko. MIR, 1981, 206 pp. Portuguese version of a 1976 Russian monograph, translated and published in Moscow. Volterra and Fredholm equations; Fourier and Laplace transform methods; approximation methods. LAS

Analysis, P. Selecta. Szolem Mandelbrojt. Gauthier-Villars, 1981, 639 pp. Reprint in a single volume of six lengthy survey articles (five in English from Rice Institute Pamphlets) on function theory, Dirichlet series, and function spaces. Concludes with a complete bibliography of Mandelbrojt's scientific publications. LAS

Analysis, P. Lecture Notes in Mathematics-889: New Classes of L^p-Spaces. Jean Bourgain. Springer-Verlag, 1981, v + 143 pp, 49.80 (P). [ISBN: 0-387-11156-5] The purpose of this text is to present new examples of L^p spaces in the sense of Lindenstrauss and Tzafriri. Some new constructions and techniques including crucial use of probabilistic results are given. PH

Algebraic Geometry, T(17-18: 1), S, P, L. Linear Algebraic Groups. T.A. Springer. Progress in Math., V. 9. Birkhäuser Boston, 1981, x + 304 pp, $18. [ISBN: 3-7643-3029-5] A basically self-contained introduction to linear algebraic groups over an algebraically closed field with minimal prerequisites from algebraic geometry. Leads to discussion of roots, Weyl group, Borel groups, existence and uniqueness theorems. Exercises, bibliography, index. JS

Algebraic Geometry, P. Lecture Notes in Mathematics-868: Surfaces Algébriques. Ed: J. Giraud, M. Raynaud. Springer-Verlag, 1981, 314 pp, $15.70 (P). [ISBN: 0-387-10842-4] Presentations from the seminars at Orsay in 1976-77 and 1977-78. JAS

Algebraic Geometry, P. The Degenerate Principal Series for Sp(2n). Robert Gustafson. Memoirs No. 248. AMS, 1981, vi + 81 pp, $4.80 (P). From the abstract: "A series of induced representations of the symplectic group of 2n x 2n matrices over a p-adic field k is decomposed." SG

Differential Geometry, P. The Spectral Theory of Toeplitz Operators. L. Boutet de Monvel, V. Guillemin. Princeton U Pr, 1981, 160 pp, $17.50; $7 (P). [ISBN: 0-691-08284-7; 0-691-08279-0]

Differential Geometry, P. Gauge-natural Bundles and Generalized Gauge Theories. David J. Eck. Memoirs No. 247. AMS, 1981, vi + 48 pp, $3.20 (P). Gauge-natural bundles, introduced in this Ph.D. dissertation, are a generalization of natural bundles which provide a natural formal context for gauge-field theory. Some development of gauge-field theory in this context is carried out. (Incidentally, this is the first mathematical paper published using AMS-TEX. It was, in fact, typeset by its author.) JAS

Geometry, T(15-16), S, L**.** Surfaces, Second Edition. H.B. Griffiths. Cambridge U Pr, 1981, xii + 128 pp, $29.95; $12.95 (P). [ISBN: 0-521-23570-7; 0-521-299772] This Second Edition differs from the First Edition (TR, April 1976) primarily in the explanation for identifying a twisted bridge. This remains a unique book of value to students who would know more of geometry for pedagogical or cultural reasons. JAS

Geometry, P. Finite Geometries and Designs. Ed: P.J. Cameron, J.W.P. Hirschfeld, D.R. Hughes. London Math. Soc. Lect. Note Ser., No. 49. Cambridge U Pr, 1981, 371 pp, $34.95 (P). [ISBN: 0-521-283787] The proceedings of the conference held June 15-19, 1980 at the White House Conference Center of the University of Sussex, England. JAS

Algebraic Topology, P. Categorical Framework for the Study of Singular Spaces. William Fulton, Robert MacPherson. Memoirs No. 243. AMS, 1981, vi + 165 pp, $9.60 (P). The author introduces a "new formalism called bivariant theories" which includes both homology and cohomology with their products and natural transformations as special cases. This machinery is then used to obtain a generalized Riemann-Roch theorem. The material grew out of work with singular spaces. JAS

Operations Research, S(15-16), P. Case Studies in Mathematical Modeling. Ed: William E. Boyce. Pitman Pub, 1981, xiii + 386 pp, $39.95. [ISBN: 0-273-08486-0] An attractive compilation of seven projects described by experienced practitioners in a chatty style. This should relieve the anxiety of some who are reluctant to try practical applications. Accessible to upper division undergraduates. AWR

Operations Research, S(16), P. Redundancy and Linear Programs. J. Telgen. Math. Centre Tracts, V. 137. Math Centrum, 1981, iii + 125 pp, Dfl. 15,75 (P). [ISBN: 90-6196-215-3] Addresses the question of how a linear programming problem should be formulated, stressing the importance of the quality of

one's model and the value of a minimal representation of the problem. The second part of the book deals with computational complexity. AWR

Operations Research, S(16-18), P, L. Operations Research: Mathematics and Models. Ed: Saul I. Gass. Proc. of Symposia in Appl. Math., V. 25. AMS, 1981, x + 198 pp, $15.60. [ISBN: 0-8218-0029-9] Six mathematical models--warfare, queueing networks, fishery management, health care delivery systems, agriculture, fire companies--from the August 1979 AMS short course in Duluth, Minnesota. LAS

Operations Research, S(16-18), P, L*. Game Theory and its Applications. Ed: William F. Lucas. Proc. of Symposia in Appl. Math., V. 24. AMS, 1981, viii + 125 pp, $18. [ISBN: 0-8218-0025-6] Lecture notes from the January 1979 AMS Short Course in Biloxi, Mississippi. Topics: cooperative games, non-cooperative games, valuation and measurement of power, and economic market games. Could serve well as the basis for an advanced undergraduate seminar. LAS

Optimization, T(17). The Theory of Games and Markets. J. Rosenmüller. Elsevier North Holland, 1981, viii + 554 pp, $73.25. [ISBN: 0-444-85482-7] Mathematically sophisticated text "for students...adapted to the scheme of definitions, theorems, and proofs. ...Economic arguments...are not very elaborate." The main emphasis is on game theory with finitely many players. Based on a four semester course at Karlsruhe University. Writing will offend if not confuse those who care about the English language. AWR

Probability, P. Lecture Notes in Mathematics-851: Stochastic Integrals. Ed: D. Williams. Springer-Verlag, 1981, ix + 540 pp, $27 (P). [ISBN: 0-387-10690-1] Proceedings of a symposium held by the London Mathematical Society and University of Durham, July 7-17, 1980. JAS

Probability, P. Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Marcel F. Neuts. Ser. in Math. Sci., V. 2. Johns Hopkins U Pr, 1981, xiii + 332 pp, $32.50. [ISBN: 0-8018-2560-1] Steady state solutions to Markov processes. Phase type distributions. Block tri-diagonal transition probability matrices. Applications include the GI/PH/1 queue. RWN

Probability, P. Lecture Notes in Statistics-5: Stationary Random Processes Associated with Point Processes. Tomasz Rolski. Springer-Verlag, 1981, vi + 139 pp, $12 (P). [ISBN: 0-387-90575-8] Technical monograph dealing primarily with continuous time theory. Concludes with a chapter giving applications of the theory to some non-standard single server queues. RSK

Probability, P. Essentials of Brownian Motion and Diffusion. Frank B. Knight. Math. Surveys, No. 18. AMS, 1981, xiii + 201 pp, $34.40. [ISBN: 0-8218-1518-0] Written to make the fundamental theory more accessible to the nonspecialist having minimal familiarity with continuous time stochastic processes. Presents those results having "immediate observational meaning, or which seem to contribute most to a general understanding of the subject." RSK

Probability, P. Probability Distributions on Linear Spaces. N.N. Vakhania. Prob. and Appl. Math. Elsevier North Holland, 1981, xiv + 123 pp, $34.95. [ISBN: 0-444-00577-3] Approaches the theory of probability distributions on Hilbert and Banach spaces from a thorough treatment of characteristic functions and Gaussian distributions on R^N and ℓ_p. TAV

Probability, T(18: 2), P. A Course in the Theory of Stochastic Processes. A.D. Wentzell. Transl: S. Chomet. McGraw-Hill, 1981, x + 304 pp, $39.95. [ISBN: 0-07-069305-6] Written as a text, with numerous problems. The author assumes measure and integration as well as functional analysis. The emphasis is on theory, beginning with the stochastic calculus, and emphasizing the Markovian theory and diffusion. TAV

Probability, P. Lecture Notes in Mathematics-863: Processus Aléatoires à Deux Indices. Ed: H. Korezlioglu, G. Mazziotto, J. Szpirglas. Springer-Verlag, 1981, 274 pp, $13.20 (P). [ISBN: 0-387-10832-7] Proceedings of a conference held June 30 and July 1, 1980 at Paris with the support of l'École Nationale Supérieure des Télécommunications, and the Centre National d'Études des Télécommunications. JAS

Probability, P. Stochastic Processes and Integration. M.M. Rao. Sijthoff & Noordhoff, 1979, xi + 456 pp, $55. [ISBN: 90-286-0438-3] An extensive treatment of the basic theoretical questions in the theory of stochastic processes including the questions of existence of consistant families. A comprehensive development of martingales and submartingales is followed by the theory of stochastic integration. TAV

Probability, T(18: 2), P. Stochastic Differential Equations and Diffusion Processes. Nobuyuki Ikeda, Shinzo Watanabe. Math. Lib., V. 24. Elsevier North-Holland, 1981, xiv + 464 pp, $85.25. [ISBN: 0-444-86172-6] The authors attempt to provide a self-contained treatment of the developments in the theory of stochastic calculus since Ito's introduction some 40 years ago. Special attention is paid to the application of these techniques to the important class of diffusion process. Note the price! TAV

Statistics, T*(13: 1). Elements of Statistical Inference, Fifth Edition. David V. Huntsberger, Patrick Billingsley. Allyn & Bacon, 1981, xiv + 505 pp, $15.95. [ISBN: 0-205-07305-0] Revision of the authors' 1977 Fourth Edition (TR, November 1977) with more emphasis on real world applications. Other new features include chapter summaries after each chapter, and a 65% increase in problems, now coded to indicate area of application and placed throughout the chapter instead of all at the end.

Statistics, T*(16-17: 1), S, P, L. Introduction to Multivariate Analysis. Christopher Chatfield, Alexander J. Collins. Chapman & Hall, 1980, x + 246 pp, $25. [ISBN: 0-412-16030-7] Clearly written blend of theory and practice requiring a basic background in probability theory, statistics and matrix algebra. Emphasizes such techniques as preliminary data analysis, principal component analysis, and procedures based on multivariate normal theory, including multivariate analysis of variance and covariance, multidimensional scaling, and cluster analysis. RSK

Statistics, P. Lecture Notes in Statistics-6: Multiple Statistical Decision Theory: Recent Developments. Shanti S. Gupta, Deng-Yuan Huang. Springer-Verlag, 1981, 104 pp, $9.80 (P). [ISBN: 0-387-90572-3] Applies general decision theory to develop principles of multiple decision procedures. Emphasizes recent developments in the theory of selection and ranking. Good set of references. RSK

Statistics, P*. Finite Mixture Distributions. B.S. Everitt, D.J. Hand. Chapman & Hall, 1981, xi + 143 pp, $14.95. [ISBN: 0-412-22420-8] Concerned with statistical distributions which can be expressed as a weighted average of a finite number of simpler components, typically normal, exponential, binomial or Poisson. Deals with the problem of parameter estimation, primarily by maximum likelihood or the method of moments, when the number of components is known, with some discussion of how to estimate this number. Good set of references, most from the last 20 years. RSK

Statistics, P. Lecture Notes in Statistics-7: Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency. Masafumi Akahira, Kei Takeuchi. Springer-Verlag, 1981, 242 pp, $14.80 (P). [ISBN: 0-387-90576-6] Collection of results on higher order (second and third) asymptotic efficiency in both regular and non-regular situations. RSK

Statistics, S*(17-18), P. Statistics in the Pharmaceutical Industry. Ed: C. Ralph Buncher, Jia-Yeong Tsay. Statistics, V. 36. Dekker, 1981, x + 465 pp, $55. [ISBN: 0-8247-1163-7] Excellent collection of articles on how the pharmaceutical industry uses statistical information in the typical 7-12 year period from the creation of the chemical in a laboratory until a drug is finally marketed. Written by 26 professionals in the field, the articles range from expository to technical. Indicates the many opportunities for statisticians in this industry, which is one of the major employers of biostatisticians. RSK

Statistics, T(16-17: 1), P, L. Simulation and the Monte Carlo Method. Reuven Y. Rubinstein. Wiley, 1981, xv + 278 pp, $32.95. [ISBN: 0-471-08917-6] In addition to the standard material on the generation of random numbers and variates, variance reduction techniques, and the solution of linear and integral equations, this textbook includes chapters on regenerative simulation and Monte Carlo optimization. The bibliography is very extensive. AO

Statistics, P. Lecture Notes in Statistics-8: The First Pannonian Symposium on Mathematical Statistics. Pál Révész, Leopold Schmetterer, V.M. Zolotarev. Springer-Verlag, 1981, vi + 308 pp, $19.80 (P). [ISBN: 0-387-90583-9] Proceedings of the September 16-21, 1979 conference at Bad Tatzmannsdorf, Austria. JAS

Statistics, S(16-17). Applied Statistics: Principles and Examples. D.R. Cox, E.J. Snell. Chapman and Hall, 1981, viii + 189 pp, $14.95 (P); $32. [ISBN: 0-412-16570-8; 0-412-16560-0] Divided into two parts. Part 1 briefly discusses general issues and principles involved in the statistical analysis of data. The lengthier Part 2 contains 24 real examples illustrating these principles, using a variety of techniques, and 15 additional real data sets. RSK

Statistics, T(15: 2), L. Probabilistic Modeling and Analysis in Science and Engineering. T.T. Soong. Wiley, 1981, xiii + 384 pp, $25.95. [ISBN: 0-471-08061-6] A post-calculus treatment, about 2/3 on probability and random variables, the rest on parameter estimation and regression. The text is clearly written with numerous examples and exercises, mostly from physics and engineering. With a stated emphasis on modeling, the split between probability and statistics seems appropriate. TAV

Statistics, T(17: 1), P. The Analysis of Categorical Data, Second Edition. R.L. Plackett. Griffin's Stat. Mono. & Courses, No. 35. Macmillan Pub, 1981, xii + 207 pp, $18.95 (P). [ISBN: 0-02-850420-8] Revision of the author's 1974 monograph dealing with the analysis of data obtained by counting. Presumes a background in statistical inference. Extensive set of references. RSK

Statistics, T(17: 1), P. Classification: Methods for the Exploratory Analysis of Multivariate Data. A.D. Gordon. Mono. on Appl. Prob. and Stat. Chapman & Hall, 1981, xii + 193 pp, $26. [ISBN: 0-412-22850-5] An "introduction to, and critical assessment of, clustering and geometrical methods of analyzing multivariate data." Final chapter contains analyses of three real data sets. Good set of references. No exercises. RSK

Statistics, P*. Frontiers in Statistical Quality Control. Ed: H.-J. Lenz, G.B. Wetherill, P.-Th. Wilrich. Physica-Verlag, 1981, 294 pp, (P). [ISBN: 3-7908-0255-7] Revised versions of papers presented at an international workshop held in Berlin in June 1980. Presents frontier results in the areas of sampling inspection and process control. RSK

Statistics, T(17: 1), P*. Survival Analysis. Rupert G. Miller, Jr., Gail Gong, Alvaro Munoz. Wiley, 1981, xi + 238 pp, $18.95 (P). [ISBN: 0-471-09434-X] In the Wiley Series in Probability and Mathematical Statistics. Reproduced from typescript. Concerned with the analysis of positive-

valued random variables (e.g., time to failure) based on censored observations. Emphasizes more recent nonparametric approaches with applications to medical research. Good set of references. RSK

Statistics, S(15-17). Solutions in Statistics and Probability. Edward J. Dudewicz. Amer Sci Pr, 1980, iv + 317 pp, $24.95 (P). [ISBN: 0-935950-00-1] Essentially a detailed solutions manual for the problems in the author's calculus-based text Introduction to Statistics and Probability (TR, August-September 1976). Problems are not stated, so it cannot be used apart from the text. RSK

Computer Programming, P. Science and Engineering Programs, Apple II Edition. Ed: John Heilborn. Osborne/McGraw-Hill, 1981, 223 pp, $15.99 (P). [ISBN: 0-931900-63-2] Listing and sample runs of 46 Applesoft BASIC programs for simple data analysis, roots of polynomials, linear algebra, differential and integral equations, Fourier series, structured analysis, thermodynamics. Requires a minimum of intelligence to key in and run. Reviewer cannot vouch for the quality of the programs or the numerical methods employed. GHM

Computer Programming, S(13-15), L. The BYTE Book of Pascal, Second Edition. Ed: Blaise W. Liffick. BYTE/McGraw-Hill, 1979, vi + 333 pp, $25. [ISBN: 0-07-037823-1] Reprint of articles (comments and opinion, language details, applications) and listings (p-code interpretor, 8080 run time code, etc.) about Pascal, originally published in 1979. A good, diverse introduction. LAS

Software Systems, S(13-14), L. Nailing Jelly to a Tree. Jerry Willis, William Danley, Jr. Dilithium Pr, 1981, viii + 244 pp, $12.95 (P). [ISBN: 0-918398-42-8] A hobbyists' guide to systems software for home computers. JL

Software Systems, L. Osborne CP/M User Guide. Thom Hogan. Osborne/McGraw-Hill, 1981, x + 283 pp, $12.99 (P). [ISBN: 0-931988-44-6] A very readable and thorough user's manual with at least some inaccuracies, e.g., the implication that all TRS-80 and Heath/Zenith microcomputers use special (high origin address) versions of CP/M. Very commendable attention is given to background and comparison with other operating systems as well as to the entire physical system of disk drives, printers, terminals, and even modems. However, don't bother with this book for "deep" system information about disk structure and BDOS details. JAS

Software Systems. How to Get Started with CP/M (Control Program for Microcomputers). Carl Townsend. Dilithium Pr, 1981, 127 pp, $9.95 (P). [ISBN: 0-918398-32-0] A rather superficial introduction to using CP/M and applications programs. Contains a lot of information appropriate for the beginning hobbyist or personal computer user, although burdened with a tendency towards jargon. Very little real information on the position of CP/M relative to its competition. JAS

Software Systems, T*(15-17: 1, 2), S*, P*, L*. An Introduction to Database Systems, Third Edition. C.J. Date. Addison-Wesley, 1981, xxviii + 574 pp, $20.95. [ISBN: 0-201-14471-9] An updated and up-to-date edition of one of the best-known and finest texts on the subject. Emphasizes the relational approach and IBM database systems, as did previous editions. Highly recommended. (First Edition, TR, March 1976.) JL

Computer Science, P. Stochastic and Deterministic Averaging Processors. P. Mars, W.J. Poppelbaum. IEE, 1981, ix + 157 pp, $52.50. [ISBN: 0-906048-44-3] This book describes the design and application of processors in which an information representation is employed that requires the use of time-averaging to obtain useful information (e.g., serial synchronous digital systems). AO

Computer Science, T(15-16: 1), S, P, L*. Computer Structures: Principles and Examples. Daniel P. Siewiorek, C. Gordon Bell, Allen Newell. McGraw-Hill, 1982, xvi + 926 pp, $31. [ISBN: 0-07-057302-6] Sequel to Computer Structures: Readings and Examples, 1971, by the latter two editors. Over 50 authoritative articles on computer architecture emphasizing developments during the last decade: micro-processors, calculators, mini-computers, vector processors, computer families, language-based computers, concurrency, networks, fault tolerances. RWN

Computer Science, P. Minicomputer Research and Applications. Ed: Helen K. Brown. Pergamon Pr, 1981, xii + 384 pp, $40. [ISBN: 0-08-027567-2] Typescript proceedings of the first conference (August 1980) of the HP/1000 User's Group. LAS

Computer Science, T(17-18: 1). The Architecture of Pipelined Computers. Peter M. Kogge. McGraw-Hill, 1981, xii + 334 pp, $28. [ISBN: 0-07-035237-2] A book about the design of very high-speed computers using pipelining. Although the focus is primarily on hardware design, some attention is given to programming techniques for the efficient use of such machines. AO

Computer Science, T(17), P. Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory. Joseph E. Stoy. MIT Pr, 1981, xxx + 414 pp, $12.50 (P). [ISBN: 0-262-69076-4] Paperback edition of 1977 original edition (TR, November 1978). GHM

Control Theory, T(16-17), L. Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management, Series Volume 4. Morton I. Kamien, Nancy L. Schwartz. Elsevier North Holland, 1981, xi + 331 pp, $32.95. [ISBN: 0-444-00424-6] An effort to explain very simply the methods of dynamic optimization. Part I is traditional calculus of variations. Part II is optimal control. For the intended readers, an appendix reviews calculus, mathematical programming, and differential equations. AWR

Control Theory, T(18: 2), P. Optimal Control of Distributed Parameter Systems. N.U. Ahmed, K.L. Teo. Elsevier North Holland, 1981, xiii + 430 pp, $75. [ISBN: 0-444-00559-5] A high-powered book organizing a good deal of research material for the first time. Demanding a good mathematical background (that can deal with abstract differential equations on a Banach space, for instance), the book prepares the reader for current research problems. AWR

Control Theory, P. Algorithmic Methods in Optimal Control. W.A. Gruver, E. Sachs. Research Notes in Math., No. 47. Pitman Pub, 1980, x + 233 pp, $29.95 (P). [ISBN: 0-273-08473-9] Deals with optimal control problems that can be described by differential equations. The techniques can be viewed as extensions of nonlinear programming for the control of dynamic systems, and the reader needs background in functional analysis and differential equations. AWR

Control Theory, P. Linear Estimation and Stochastic Control. M.H.A. Davis. Chapman & Hall, 1977, xii + 224 pp, $12.95 (P); $16.95. [ISBN: 0-470-99215-8] Paperback republication of 1977 original edition (TR, April 1978). LAS

Control Theory, T(17: 1, 2), S, L. The Calculus of Variations and Optimal Control: An Introduction. George Leitmann. Math. Concepts and Methods in Sci. and Eng., V. 24. Plenum Pr, 1981, xvi + 311 pp, $35. [ISBN: 0-306-40707-8] The first quarter covers the foundations of the calculus of variations needed for the remainder of the book, which is on the optimal control of dynamical systems: necessary and sufficient conditions; isoperimetric, state-dependent control; and endpoint inequality constraints. Examples. RWN

Control Theory, P. Lectures on Stochastic Control Theory. Makiko Nisio. ISI Lect. Notes, No. 9. Macmillan India, 1981, vii + 83 pp, Rs. 25.00 (P). Lectures by Nisio given at the Indian Statistical Institute in 1978, based on the Bellman principle. High priced for a short, typewriter-to-print booklet. AWR

Control Theory, P. Lecture Notes in Control and Information Sciences-32: Realization Theory of Continuous-Time Dynamical Systems. Tsuyoshi Matsuo. Springer-Verlag, 1981, vi + 329 pp, $17.40 (P). [ISBN: 0-387-10682-0] The realization problem is to determine an intrinsic mathematical model from the input-output relations (given by experimental data) of a causal black box. This monograph presents a basis for realization theory of dynamical (as opposed to discrete) systems, starting from axioms for input experiments called a concatenation monoid. LAS

Systems Theory, P. Nonlinear Dynamics. Ed: Robert H.G. Helleman. Annals of New York Acad. of Sci., V. 357. New York Acad of Sci, 1980, xi + 507 pp, $98 (P). [ISBN: 0-89766-104-4] Proceedings of an International Conference held by the New York Academy of Sciences, December 17-21, 1979. JAS

Applications, P. The Second Workshop on Grand Unification. Ed: Jacques P. Leveille, Lawrence R. Sulak, David G. Unger. Birkhäuser Boston, 1981, ix + 321 pp, $19.95. [ISBN: 3-7643-3055-4] Proceedings of a workshop held at the University of Michigan, April 24-26, 1981. JAS

Applications (Behavioral Science), P. Man-Computer Interaction: Human Factors Aspects of Computers & People. Ed: B. Shackel. Series E: Appl. Sci., No. 44. Sijthoff & Noordhoff, 1981, x + 560 pp, $60. [ISBN: 90-286-0910-5] Very dated proceedings of a September 1976 NATO Advanced Study Institute held in Mati, Greece. (Six years to publish a typescript volume on such a rapidly changing subject is inexcusable.) Mostly devoted to behavioral studies of man-machine symbiosis. LAS

Applications (Biology), T(17: 1), P*, L. The Mathematical Theory of Quantitative Genetics. M.G. Bulmer. Clarendon Pr, 1980, x + 255 pp, $74. [ISBN: 0-19-857530-0] Presents underlying theory of quantitative genetics, "the study of the inheritance of...continuous, quantitative characters" (as opposed to the discrete "characters" of classical genetics). Presumes some knowledge of statistics, particularly regression and analysis of variance. Note price. RSK

Applications (Biology), P, L. Some Mathematical Questions in Biology. Stephen Childress. Lect. on Math. in the Life Sci., V. 14. AMS, 1981, x + 214 pp, $19.20 (P). [ISBN: 0-8218-1164-9] Six papers--three on developmental biology, three on biomechanics--from the Symposium on Mathematical Biology at the January 1981 AAAS meeting in Toronto. LAS

Applications (Classical Mechanics), P. Classical Mechanics and Dynamical Systems. Ed: Robert L. Devaney, Zbigniew H. Nitecki. Lect. Notes in Pure and Appl. Math., V. 70. Dekker, 1981, x + 237 pp, $35 (P). [ISBN: 0-8247-1529-2] Proceedings of a conference held at Tufts University in August 1979. JAS

Applications (Electronics), S(15-16). Interfacing to S-100/IEEE 696 Microcomputers. Sol Libes, Mark Garetz. Osborne/McGraw-Hill, 1981, xiv + 321 pp, $15 (P). [ISBN: 0-931988-37-3] A very detailed treatment of the S-100 bus signals with a number of detailed examples using the S-100 bus such as assorted interfacing techniques, digital to and from analog conversion, and timers and counters. JAS

Applications (Electronics), S(13-16), L. How to Build Your Own Working Microcomputer. Charles K. Adams. TAB Books, 1980, 308 pp, $9.95 (P). [ISBN: 0-8306-1200-9] A half-way point between a digital electronics text and a Heathkit construction manual, this book explains the circuitry from a user's point of view, lays out block and digital diagrams, actual component placement and wiring, and discusses in some detail programming the resulting 8080 computer. It even includes information on

interfacing and on wiring in a pocket calculator chip. Looks useful for the experienced kit builder who is involved in microcomputing and would like to get his hands dirty while learning about the heart of the matter. JAS

Applications (Engineering), T*(16-17: 1, 2), S*, P*, L*. Mathematical Theory of Wave Motion. G.R. Baldock, T. Bridgeman. Math. and its Appl. Wiley, 1981, 261 pp, $64.95. [ISBN: 0-470-27113-2] Solid, thorough, no-nonsense, example-laden study expertly written by practitioners of the discipline. Emphasis on the unity of methods and concepts applied to the wave equation and its generalizations, including treatments of dispersion and waves in discrete structures. Aimed at undergraduate students in mathematics, physics, engineering and oceanography, but will tax all but the best-prepared. Suitable for graduate students and practicing engineers. Exercises, with answers where appropriate. Good index. JK

Applications (Engineering), T(17-18: 1), S, P. Digital Foundations of Time Series Analysis, Volume 2: Wave-Equation Space-Time Processing. Enders A. Robinson, Manuel T. Silvia. Holden-Day, 1981, viii + 534 pp, $35. [ISBN: 0-8162-7271-9] This text presents analytic and numerical methods for processing empirical data associated with multidimensional models such as Poisson's equation and the wave equation. Applications of these techniques in areas such as underwater acoustics and geophysical exploration are also presented. AO

Applications (Engineering), T(16-17: 1, 2). Concepts and Applications of Finite Element Analysis, Second Edition. Robert D. Cook. Wiley, 1981, xix + 537 pp, $34.95. [ISBN: 0-471-03050-3] The finite element method is a numerical procedure for solving continuum mechanics problems. This introductory textbook emphasizes applications of the method in stress analysis and structural mechanics. AO

Applications (Engineering), S(15-16). Microprocessors for Measurement and Control. David M. Auslander, Paul Sagues. Osborne/McGraw-Hill, 1981, x + 310 pp, $15.99 (P). [ISBN: 0-931988-57-8] A rather engineering-oriented discussion of several projects such as DC Motor Control and Testing, and Automatic Weighing. Provides discussion of 8080, 8085, Z80, and PDP-11 programs in various languages. JAS

Applications (Engineering), S(15-16), P. Case Studies in Mathematical Modelling: A Course Book for Scientists and Engineers. Ed: R. Bradley, R.D. Gibson, M. Cross. Halsted Pr, 1981, 167 pp, $37.95. [ISBN: 0-470-27235-X] Intended as a text for recent recruits to industry or other persons interested in increasing their understanding of modelling. The work of six groups on six different problems is described in a way that exhibits group inclinations before giving the model preferred by the experienced group supervisors. AWR

Applications (Engineering), P. Problems of Elastic Stability and Vibrations. Ed: Vadim Komkov. Contemporary Math., V. 4. AMS, 1981, x + 137 pp, $8.40 (P). [ISBN: 0-8218-5005-9] Enlarged versions of talks presented at the May 1981 Pittsburgh meeting of the AMS. Emphasizes problems of near-singularity and multiple eigenvalues. LAS

Applications (Fluid Dynamics), P. Transition and Turbulence. Ed: Richard E. Meyer. Academic Pr, 1981, ix + 245 pp, $15.50. [ISBN: 0-12-493240-1] Twelve papers from an October 1980 symposium at the Mathematics Research Center in Madison, Wisconsin concerning the importance of the relation between transition and turbulence for understanding of real fluid motion. LAS

Applications (Geophysics), T(17-18: 1), S, P. An Introduction to the Mathematical Theory of Geophysical Fluid Dynamics. Susan Friedlander. Math. Stud., V. 41. Elsevier North Holland, 1980, x + 272 pp, $29.50 (P). [ISBN: 0-444-86032-0] A text intended for graduate students in applied mathematics. It presents a mathematical description of problems in fluid dynamics for which the length scale is sufficiently large that the earth's rotation has a significant effect. AO

Applications (Information Theory), L. An Introduction to Information Theory: Symbols, Signals & Noise, Second, Revised Edition. John R. Pierce. Dover, 1980, xii + 305 pp, $4.50 (P). [ISBN: 0-486-24061-4] Updated, revised edition of Symbols, Signals and Noise: The Nature and Process of Communication (Harper & Row, 1961). Account for nonmathematical layman of Shannon's theory of communication. The mathematics is faced squarely but in the most elementary terms possible, with a minimum of notation and a glossary and appendix to help out. Discusses applications to linguistics, coding, art, psychology and physics. GHM

Applications (Linguistics), S(18), P. Pratique de L'Analyse des Données, T. 3: Linguistique & Lexicologie. J.-P. Benzécri, et al. Dunod, 1981, x + 565 pp, (P). [ISBN: 2-04-010776-2] A collection of papers on the statistical analysis of linguistic data. JD-B

Applications (Physics), S*(15-17), P, L. To Fulfill a Vision: Jerusalem Einstein Centennial Symposium on Gauge Theories and Unification of Physical Forces. Ed: Yuval Ne'eman. Addison-Wesley, 1981, xxxi + 279 pp, $39.50. [ISBN: 0-201-05289-X] This volume from the March 1979 Symposium in Jerusalem represents a mildly technical expository treatment of today's particle physics. It is logically paired with the earlier volume Some Strangeness in the Proportion which reports on today's relativity theory as presented in a parallel Princeton symposium in March 1979. The treatment is very current, given by great experts, yet conveys an unusual sense of history and of the humanity of Einstein and his predecessors. JAS

Applications (Physics), T(17-18), S, P. Gauge Fields: Introduction to Quantum Theory. L.D. Faddeev, A.A. Slavnov. Benjamin/Cummings Pub, 1980, xiii + 232 pp, $28.50. [ISBN: 0-8053-9016-2] Volume 50 in the Frontiers in Physics Series, translated from the 1978 Russian edition, this is an expository introduction to quantum dynamics of gauge fields. It requires at least an advanced undergraduate background in physics or somewhat less physics and a good background in the differential geometry of modern physics. JAS

Applications (Physics), S(16-18), P, L. Introduction to Tensors, Spinors, and Relativistic Wave-Equations (Relation Structure), Second (Unaltered) Edition. E.M. Corson. Chelsea Pub, xii + 221 pp, $14.95. [ISBN: 0-8284-0315-5] Reprint on long-life paper of a work first published in Glasgow in 1953. A classical study of covariance, group theory, space-time and field theory to explain how and why particular formalisms are suitable for the description of particles of given spin. LAS

Applications (Social Science), S*(17), P*, L. Stochastic Models for Social Processes, Second Edition. D.J. Bartholomew. Wiley, 1978, xi + 411 pp, $57.50. [ISBN: 0-471-05451-8] Reprint of the 1973 Second Edition (TR, November 1974). Everything is the same but the price--$19.50 then!. RSK

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

New Jersey Section

The fall meeting of the New Jersey section was held on October 24, 1981 at Trenton State College in Trenton, New Jersey. Approximately fifty members of the Association were in attendance.

Invited Lectures:

"Maps from the Interval to Itself," by John Milnor, Institute for Advanced Study.
"American Women in Mathematics: The Pre-World War II Ph.D.'s," by Judy Green, Rutgers University.
"Mathematical Models and Existence Theorems," by Dorothy Bernstein, Brown University.

Panel Discussion:

"What Should Employers Expect from Mathematical Sciences Graduates?" chaired by Edward Conjura, Trenton State College. Panel members were George Byrne, Exxon Research and Engineering Company; Erna Hoover, Bell Laboratories; Thomas Spencer III, Analytical Support for Bell of Pennsylvania; Brenda Wilson, Academy of Natural Sciences in Philadelphia; Robert Urbanski, Middlesex County College.

Northeastern Section

The annual meeting of the Northeastern section was held at Trinity College in Hartford, Connecticut, November 20-21, 1981. Approximately 64 people attended the meeting.

Invited Lectures:

"The Biography of a Theorem," by Dorothy L. Bernstein, Brown University.
"Number Theory on Elliptic Curves--Old Theorems and Recent Conjectures," by John T. Tate, Harvard University. (Christie Lecture)
"Curves of Constant Width," by Bruce B. Peterson, Middlebury College.

Panel Discussion:

"Mathematics Curricula for Non-Mathematics Majors," chaired by Carroll F. McMahon, Bentley College. Panel members were Karen Schroeder, Bentley College; Linda Wilkins, Regis College; Donald Small, Colby College.

That is, the graph of $h_n(x)$ coincides with that of $y = x^2$ at the end points of the interval $[1/2^n, 1/2^{n-1}]$, has slope 1 at each end point, and has a point of inflection at the midpoint of the interval, where the slope is 0. Now, for $x \geqslant 0$, define

$$h(x) = \begin{cases} 0 & \text{if } x = 0 \\ h_n(x) & \text{if } 1/2^n \leqslant x < 1/2^{n-1}, \quad n = 2, 3, \ldots \\ x - \dfrac{1}{4} & \text{if } x \geqslant \dfrac{1}{2} \end{cases}$$

and for $x < 0$ define $h(x) = -h(-x)$. Restricting our attention to positive x since the curve is symmetric in the origin, we have, for $n = 2, 3, \ldots,$

$$h(1/2^n) = 1/2^{2n}, \quad h'(1/2^n) = 1, \quad h'(3/2^{n+1}) = 0.$$

For the derivative h', $h'(x) \geqslant 0$, and, in fact, h is a strictly increasing function. Since $h'(x)$ takes on all values in $[0, 1]$ in every interval $(0, \delta)$, the derivative $h'(x)$ fails to be continuous at the origin. It is easy to verify that $h'(0) = 0$. Finally, the function

$$k(x) = \begin{cases} -x^2 + h(x) & \text{if } x < 0 \\ x^2 + h(x) & \text{if } x \geqslant 0 \end{cases}$$

has all the desired properties.

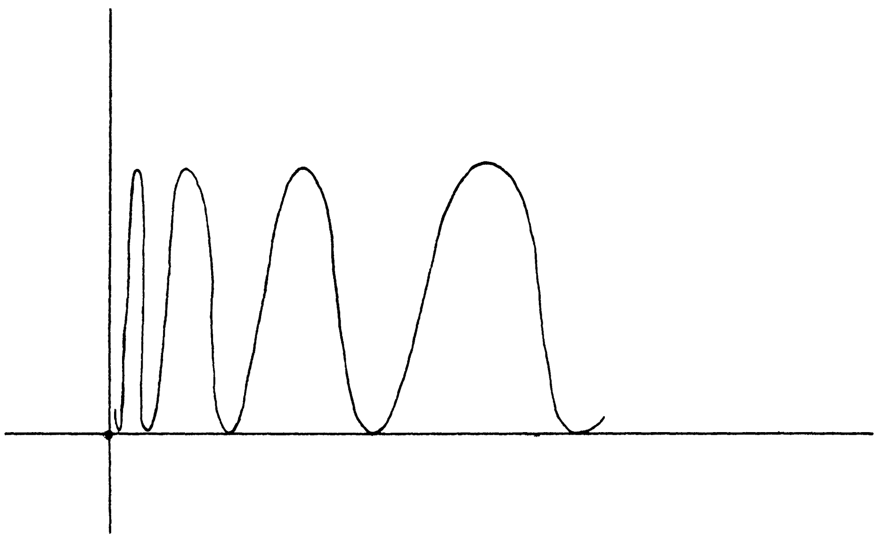


FIG. 3. $h'(x)$ $(h'(0) = 0)$

Reference

1. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.

MISCELLANEA

70. Everybody praises the incomparable power of the mathematical method, but everybody is also aware of its incomparable unpopularity.

—J. Rosanes, *Jber. Deutsch. Math. Verein.*, 13 (1904) 17.

$$\int_a^b k_x(t,x)P(x)\,dx = P(c)\int_a^b k_x(t,x)\,dx = P(c)[k(t,b)-k(t,a)],$$

where $a < c < b$. Hence

$$\int_a^b k(t,x)p(x)\,dx = k(t,b)[P(b)-P(c)] + k(t,a)[P(c)-P(a)],$$

$$\left|\int_a^b k(t,x)p(x)\,dx\right| \leq M|P(b)-P(c)| + M|P(c)-P(a)|$$

$$= M\left|\int_c^b p(x)\,dx\right| + M\left|\int_a^c p(x)\,dx\right|.$$

Since $p(x)$ is integrable, we can make the integrals on the right as small as we please (independent of t) by choosing a sufficiently large. Hence the convergence is uniform in t.

References

1. R. Courant, Differential and Integral Calculus II, Wiley, New York, 1936, p. 317.
2. J. Edwards, Treatise on Integral Calculus II, Chelsea, New York, 1922, p. 325.
3. G. M. Fikhtengol'ts, Fundamentals of Mathematical Analysis II, Pergamon, London, 1965, p. 151.
4. C. Jordan, Cours d'Analyse II, 2/e, Gauthier-Villars, Paris, 1893, p. 199.
5. P. Franklin, Methods of Advanced Calculus, McGraw-Hill, New York, 1944, p. 270.
6. H. S. W. Massey and H. Kestelman, Ancillary Mathematics, 2/e, Pitman & Sons, London, 1964, p. 450.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

MINIMAL MATHEMATICAL COMPETENCIES FOR COLLEGE GRADUATES*

1. Introduction. Too many people know too little mathematics. Even those who are well informed in other ways often cannot appreciate, much less participate in, some major currents of modern life because of their ideas and feelings about mathematics. In a relatively severe but all too common form, ignorance of mathematics amounts to a form of "functional illiteracy."

Along with the recent revival of interest in general education, "core" curricula, and minimal competencies, this problem has naturally led to the question: What mathematics should every graduate of an American college or university know?

At its January, 1978, meeting, the Association's Committee on the Undergraduate Program in Mathematics (CUPM) established a panel to study the question and make appropriate recommendations. The members of the panel are: Gerald L. Alexanderson (University of Santa Clara), Robert J. Bumcrot (Hofstra University), Edwin H. Spanier (University of California, Berkeley), Juanita J. Peterson (Laney College), and Donald W. Bushaw (Washington State University; chairman).

Some of the work of the panel is described in an Appendix to this document, which is a report from the panel.

---

*Report of a CUPM Panel whose members are listed in the third paragraph. The report has been approved by CUPM itself.

2. Recommendations. The leading lesson the panel learned from its surveys (see the Appendix) is that American colleges and universities are so diverse that it is impossible to describe either an approximately standard practice or an everywhere attainable goal. A set of minimal competencies that might be woefully inadequate for specialized or selective universities can be a hopeless ideal for others. To perform its task realistically, the panel has therefore felt obliged to interpret the word "minimal" in a really minimal way. *The recommendations listed below accordingly refer to a bare minimum of mathematical competencies for all college graduates.* The panel hopes that individual institutions will go as far beyond these recommendations as local conditions allow. Similarly, how the requirements should be met is left open, for that depends not only on the requirements themselves but also on local policies, traditions, and resources.

The following recommendations result from the panel's studies and deliberations. In preliminary form, they have been reviewed by numerous mathematicians and nonmathematicians, and have been considerably modified in light of comments received. In this sense they represent the collective judgment of a group much larger than the panel itself.

Recommendation A. All college graduates, with rare exceptions, should be expected to have demonstrated reasonable proficiency in the mathematical sciences. Every college or university should therefore formulate, with adequate concreteness, what this "reasonable proficiency" should mean for its students; define how students should demonstrate this proficiency; and establish this demonstration as a degree requirement.

Competence in arithmetic and some facility in making applications in everyday life might be a reasonable graduation requirement for two-year college students in terminal and vocational programs.

Four-year colleges and universities should normally require—perhaps on entrance—not only these but elementary algebra and elementary geometry. They should also expect graduates to understand and be able to use some elementary statistical ideas, to be aware of the place of mathematics in society generally, and to appreciate the nature and societal significance of computing. This applies also to two-year college students in university parallel curricula.

Recommendation B. Whether or not stipulated proficiency is tested by examination, courses should be made available in which it may be acquired. These courses should be taught by effective instructors, and should be designed to be appealing and significant to the students.

Recommendation C. In particular, one or more courses of a remedial nature should be available where there is a need. Such courses, by definition, ordinarily present precollege material, but it should be presented in a way suited to the clientele. In institutions where it is considered improper or impossible to offer remedial courses, mastery of the mathematics should be assured either by entrance requirements or by referring students to other schools where remedial courses can be taken. Two-year colleges have made a large contribution in this role and may be expected to continue to do so.

Is college credit appropriate for remedial courses? On this point we will only quote the statement approved by the MAA Board of Governors on August 20, 1979: "College credit granted for work in mathematics must be carefully controlled. It should not be granted for distinctly high school level work. Mathematics courses offered in college should be examined to determine the extent of their overlap with high school mathematics, and where that overlap is substantial the course should not provide credit toward college graduation; but the students should be graded on their work, and the results should be included in computing grade point averages."

Recommendation D. While almost all undergraduate courses in mathematics should give attention to applications and to historical and philosophical aspects of the subject, there should be one or more courses that concentrate on these aspects while remaining accessible to students with little mathematical background.

Recommendation E. Individual interests often lead students to take a considerable amount of post-secondary mathematics in conventional courses. These students should also be able to

take a course of the kind described in Recommendation D, but presupposing more mathematical background.

The MAA Committee on Improving Remediation Efforts in the Colleges, chaired by Professor Joan Leitzel, has gathered information about the effective remedial programs and has made its own recommendations. A separate CUPM panel, chaired by Professor Jerome Goldstein, is at the same time formulating recommendations on "mathematics appreciation" courses of the kind described in recommendations D and E and in Section 4 below. The Minimal Competencies Panel has worked in liaison with both groups and sees no conflict among the various recommendations.

Nevertheless, each of these two main matters will be discussed further in the remaining sections of this report. These discussions are intended primarily to clarify the panel's recommendations, but partly as a way of passing along some of the good ideas it has collected. The separation of the two matters is certainly not intended to imply that remedial courses should do nothing to convey an appreciation of mathematics, or that techniques are out of place in mathematics appreciation courses.

3. Mathematics for coping with life. The idea that all college graduates should be expected to have acquired a certain familiarity with mathematics rests in part on the well-founded belief that such a familiarity is necessary for effective functioning in contemporary life, and certainly for life in those spheres college graduates are most likely to enter. Indeed, it may be argued convincingly—and has been argued many times—that a modest acquaintance with mathematics is necessary for the successful functioning of almost *any* member of modern society. But any prerequisite for contemporary life in general ought to be, *a fortiori*, something one has a right to expect of all college graduates.

Unfortunately many students manage to enter college without having learned the mathematics needed for coping with everyday life, and a deplorable fraction of them leave college in the same condition. The panel's recommendations—most explicitly Recommendation C—suggest that for such students there should be at least one course where basic mathematical deficiencies may be repaired.

Students entering college with mathematical deficiencies have presumably had opportunities to learn the mathematics, and for them those opportunities did not work. Therefore, *the college remedial course should not be a mere rehash, and certainly not an accelerated one, of the traditional secondary or even primary course*. Courses that cover the same old ground in much the same old way tend to be just as uninspiring and unintelligible for these students as the originals, and therefore even less likely to succeed. Students should be able to find even remedial courses fresh, interesting, and significant.

Many courses of this type are being offered, and new ideas are being tested all the time. Several approaches have been described in print (see, for instance, the CUPM booklet *A Course in Basic Mathematics for Colleges*, reprinted in *A Compendium of CUPM Recommendations*, 1, 256-313), and other reports will surely appear. Here there will be only a sketch to illustrate the type of course that might be considered.

The goals of the course would be to impart mathematical knowledge needed for dealing with most common situations in which deductive reasoning or calculation is needed, and to provide some motivation and preparation for a second course in mathematics that could help the students become educated men and women. It is *not* a goal of the course to teach, once and for all, high school mathematics in its entirety, or to provide background for some standard courses in mathematics or other scientific subjects. (The problem of preparing students for mathematics courses required in their fields is discussed at length in the report of the Committee on Improving Remediation Efforts in the Colleges.)

Students in the course would typically have studied no mathematics for three or four years, and have been bored, mystified, or discouraged by past experiences with mathematics courses. They should be in their *first two* years of college. There should be *no* formal prerequisites.

The course should be relatively brief (twenty to thirty meetings), and should be managed in such a way that students participate actively and receive frequent personal attention. To facilitate this, there should be approximately a fifth as many student assistants as there are students. The first few times the course is offered, the assistants might be mathematics or science majors; later, they should be students who have succeeded in this and at least one further mathematics course.

Equipment might include identical calculators for the students, the assistants, and the instructor. The calculator should have the four basic arithmetical operations, sign changes, squares, square roots, floating decimal, a one-word memory, and very little else. A device for projecting the face of the instructor's calculator on a screen would be useful. There should also be a large collection of advertisements, newspaper and magazine articles, sales and credit agreements, and so on, the interpretation or use of which would require some of the topics listed below. These might be complemented by *reasonable* imaginary examples, but the illustration of no topic should depend entirely on artificial applications. If no genuine examples can be found, why should the topic be included? In some topics, however, a step should be taken beyond the evidently practical.

Students should be supplied with a single page of formulas, sufficient for the whole course.

The grading policy should be compassionate but firm. Tests should be frequent and repeatable at least once. They should be straightforward, but only high scores should be considered passing. Mastery should be recognized irrespective of the number of attempts needed to show it, within limits, but outstanding performance should be recognized. If possible, permanent records of students who need to repeat the course should not show the unsuccessful tries.

One list of topics for such a course is given below. Additions and modifications should be made in response to real-world needs and experience in offering the course.

1. Positive decimals; conversion of fractions to decimals with the calculator.
2. Pencil and paper arithmetic with signed whole numbers.
3. Pencil and paper arithmetic with signed fractions. (There should be no three-or-more digit numerators or denominators, except powers of ten.)
4. Calculator arithmetic with signed decimals.
5. Rounding off.
6. Estimation; orders of magnitude.
7. Scientific notation.
8. Units of measurement; elements of the metric system.
9. Percent.
10. What is a formula? What is a function?
11. Times, distance, and rates.
12. Area and volume.
13 What is an algorithm? Flowcharting.
14. Statistics and its dangers.
15. When is an argument correct?
16. Compound interest.
17. Exponential change.

This list should not give rise to hideous visions of workbooks filled with drill exercises. Games, problems of obvious everyday interest, opportunities for creativity, and occasional attention to general problem-solving strategies should contribute to a cheerful and progressive atmosphere and a positive experience.

4. Mathematics appreciation. While the panel does not insist that a knowledge of the cultural side of mathematics should be required of all college students, its Recommendations D and E above suggest that attractive and accessible courses dealing especially with that aspect should be offered. This section of the report contains some reasons for this position and some comments on how it might be realized.

Mathematics has played a central role in the development of modern civilization. It has been

essential not only to the growth of science and technology, but has had profound effects on philosophy and other forms of thought as well.

There was certainly no doubt in past centuries that every college graduate, to be an educated person, had to know some mathematics. In medieval times, for example, four of the seven traditional liberal arts were largely or wholly mathematical. The importance attached to mathematics was evident in courses of study in the nineteenth century, and this carried over into the twentieth. Now, however, it is possible to graduate from many colleges without any contact with mathematics beyond the most elementary high-school courses.

While high-school mathematics is important, it does tend to emphasize development of skills. The same, unfortunately, may be said of most college courses whose mission is primarily remedial or preprofessional. But an educated, well-informed person should know something about mathematics beyond skills.

To many, the distinction between mathematicians and accountants is not clear. People who are alert and informed about many things, even colleagues in a university, sometimes assume that mathematicians are constantly doing arithmetic and are surprised to hear that there is such a thing as mathematical research. Their experiences with school mathematics left them with the impression that mathematics is ancient and immutable, and consists of rules and formulas for unfortunate school children to memorize.

The great mathematicians do not occupy their rightful place in the public consciousness. In his *New Yorker* article on mathematics (February 19, 1972), Alfred Adler rightly observed that " ...it would be astonishing if the reader could identify more than two of the following names: Gauss, Cauchy, Euler, Hilbert, Riemann. It would be equally astonishing if he should be unfamiliar with the names of Mann, Stravinsky, de Kooning, Pasteur, John Dewey. The point is not that the first five are the mathematical equivalents of the second five. They are not. They are the mathematical equivalents of Tolstoy, Beethoven, Rembrandt, Darwin, Freud. The geometry of relativity—the work of Riemann—has had consequences as profound as psychoanalysis has..."

Many college graduates know a great deal of mathematics; most of them have had to take mathematics in preparation for their work. But how many of these, or how many mathematics majors, for that matter, could tell much about Abel or Jacobi? More important, how many of them could comment plausibly on the relation of mathematics to other disciplines?

The point here is not that mathematics and mathematicians should be glorified but that a reasonable perspective on the place of mathematics in the human enterprise should be more widely shared.

A course designed specifically to improve this perspective would ideally give some idea of what sorts of problems mathematicians consider and how such problems are attacked. The object would be to promote mathematical literacy, interpreted to include an awareness among future colleagues in colleges and universities, in business, in industry, in government, and in many other callings of what mathematics is, why it is important, and how it might serve them. Some history should be covered along the way, but a straight course in the history of mathematics is not recommended for this purpose; it can have meaning only if the students already have some understanding of the mathematical ideas whose development is traced.

The course could include, for example, a discussion of the Euler formula for polyhedra—and the names of Euler, Descartes, and Cauchy already would have entered the discussion. An account of non-Euclidean geometry would be appropriate, and provide an occasion for introducing Gauss and Riemann as well as Bolyai and Lobachevski, and for commenting on the element of arbitrariness in mathematical modeling of reality. Neither of these topics requires any high level of algebraic skill. A discussion of the insolubility of the quintic equation might involve more algebra but would refer to the work of Lagrange, Galois, and Abel—and the important idea of mathematical impossibility would have arisen. There are many other topics that bring up important mathematical ideas and events but do not require much background.

Axiomatics, though obviously important, should not be overemphasized. Axiomatic systems

should not be presented in detail unless one obtains by their use some interesting results that were not intuitively obvious from the start. Elementary graph theory offers some nice opportunities here, as well as a great variety of easily understood applications. Laborious efforts to prove the obvious can convince people that the whole endeavor is silly.

Applications are appealing to many students and should be included. There are convenient sources of authentic applications of mathematics at every level of difficulty. Applications, however, should not be allowed to upstage the real star of the show, mathematical thought itself. Calculators and computing might have their place in the course, and some time could profitably be spent on the place of computers in modern society. Serious study of computer science, however, is probably best left to other courses.

The course should give students copious evidence that mathematics has not only played a great part in human history, but continues to thrive in the service of other fields and as an independent source of intellectual excitement and aesthetic appeal. Mathematical "current events," such as the solution of the four color problem and the discovery of new large primes should be mentioned. Something might be said about Hilbert's problems and the Fields medals. Carefully selected readings from *Scientific American*, *The Mathematical Intelligencer*, and similar publications can help.

The choice of faculty for an appreciation course is critical. It is an extraordinary teaching assistant who would have the experience and breadth of outlook to teach such a course. It should usually be taught by senior faculty, and if appropriate faculty cannot be found, the course should not be taught at all. And it is better that it be taught by the right faculty in larger sections than by reluctant or inept instructors in small ones.

The course mentioned in Recommendation E offers further opportunities. It is still too easy for mathematics and science majors to complete their programs without knowing that research is done in mathematics, that mathematics has deep and productive relationships with many fields, and that mathematics has a rich and fascinating history. A mathematics appreciation course for students with good technical proficiency in mathematics can do much to take care of this and be a memorable experience for all concerned.

As has already been said in Recommendation D, *these observations about separate mathematics appreciation courses should apply, to some extent, to all mathematics instruction*, even remedial. In a perfect world every mathematics course would be a mathematics appreciation course. The world, however, is not perfect.

5. Conclusion. The recommendations and other ideas set forth in this report will surely not be the last word on the subject. Many intelligent people will be giving further thought to it, and future experience should certainly be allowed and expected to affect our outlook on the whole matter. Accordingly, the panel extends a standing invitation for comments on this report or on the entire question. They should be addressed to the panel in care of Prof. D. W. Bushaw, Dept. of Mathematics, Washington State University, Pullman, Washington 99164.

Appendix. *What the panel did.* The panel began by consulting the pertinent literature; officers of organizations represented in the Council of Scientific Society Presidents or the Conference Board of Mathematical Sciences; and a sample of mathematicians drawn at random from the 1978–1979 *Combined Membership List*. Summaries of the results may be obtained from the chairman of the panel.

A general announcement and appeal for information and ideas also appeared in *Notices of the American Mathematical Society*, *Change Magazine*, *The Mathematics Teacher*, *The Chronicle of Higher Education*, *The Two-Year College Mathematics Journal*, *SIAM News*, and this MONTHLY.

From the first two surveys mentioned, the panel learned not much more than that no national organization in this country, the MAA itself not excepted, has ever taken a position on what college graduates *in general* should know of mathematics.

The survey based on the *Combined Membership List* (CML) and the appeal in periodicals,

though more productive, did not provide as much unambiguous guidance as the panel had hoped to get. The CML survey yielded 335 usable responses from a thousand questionnaires. 226 were from persons at colleges and universities. Of these, 105 (39.5%) were from institutions where a mathematics requirement for graduation was in force. These 105 respondents were asked about the nature of the requirement, whether they favored it, and whether they thought it was effective. In the great majority of cases (91 or 86.7%) the requirement could be satisfied by one or more courses. Seven of these respondents reported that the requirement could be satisfied by examination; five others said both courses and an examination were required.

One hundred (95.2%) of the 105 said they favored the requirement, and 75 (71.4%) said they thought it was at least partially effective.

The median course requirement, where one existed, was between 3 and 4 semester hours. A specific course or sequence of courses was seldom required; indeed, acceptable courses were remarkably diverse.

The 161 respondents in colleges and universities which had no general mathematics requirement were asked whether they favored such a requirement. In reply, 148 expressed a preference, and of these 104 (70.3%) favored some kind of a requirement.

When the two groups are combined, one finds that 204 of 253 (80.6%) of those college- or university-affiliated mathematicians in the sample who expressed any preference favored some general graduation requirement in mathematics. The panel did not expect this fraction to be so high. (Unfortunately, the questionnaire did not ask for reasons for the preferences expressed.)

All respondents, academic or not, were asked to mark in a forty-item list of mathematical topics those they thought should be required of all college graduates. The following topics were marked by at least half of the respondents:

basic arithmetic skills (94.6%)
area and volume of common figures (76.4%)
linear equations (71.3%)
algebraic manipulations (63%)
elementary statistics (55.5%)
graphing of elementary functions (54.9%)
integer and fractional exponents (54.3%)
elementary plane geometry (51.9%)

Next in order were: elementary probability (49%), general problem-solving skills (heuristic) (49%), quadratic equations (47.5%), mathematics in business (46.9%), and radicals (43.9%). Computer programming was marked by 33.1%, just after elementary logic (35.5%) and systems of equations (35.2%).

The question about what standard courses should be required elicited a wide variety of answers, many of which were in fact far from standard. College algebra (mentioned by 51 respondents) led the list, and was followed by probability and statistics (47), calculus (45), elementary or intermediate algebra (44), and computer programming or appreciation (30).

About 45% of the respondents accepted an invitation to comment further. Many merely expanded on earlier answers, but some submitted careful statements of their views. These statements, though not easy to summarize, were carefully studied by the panel.

Responses to the appeal in periodicals were interesting too, but they are even less reducible to a brief summary.

The panel met three times and also conducted a voluminous correspondence within itself and with others. It completes this report with high respect for the complexity of the problem, but hopes that its proposals will be of some use in finding solutions.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA (USA) 94303, by August 31, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2938*. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

Can one find n points in the plane (no three on a line, no four on a circle) so that for every i, $i = 1, 2, \ldots, n-1$, there is a distance determined by these points that occurs exactly i times? (It is true for $n = 3, 4, 5, 6$.)

E 2939. *Proposed by Tian Jinghuang, Szechwan University, China.*

Consider the autonomous quadratic first-order system

$$dx/dt = X(x,y), \quad dy/dt = Y(x,y), \tag{1}$$

where

$$X(x,y) = \sum a_{ik} x^i y^k, \quad Y(x,y) = \sum b_{ik} x^i y^k$$

are relatively prime polynomials of degree at most 2 and not both linear. Let the curve l be a branch of the (possibly degenerate) conic section $X(x,y) = 0$. Let A and B be two adjacent equilibrium points of (1) lying on l. If l is one of two intersecting lines forming the locus $X(x,y) = 0$, then ind $A = \pm$ind B according as the intersection of the two lines does or does not lie between A and B. If $X(x,y) = 0$ does not consist of two intersecting lines, then ind $A = -$ind B. (Here ind A, ind B are the indices of the equilibrium points.) Prove these assertions.

E 2940. *Proposed by Andrew Leonard, Indiana University.*

If $f(x,y)$ is defined for all real x, y, and if f is a polynomial in x for every y, and a polynomial in y for every x, prove that f is a polynomial in two variables.

E 2941. *Proposed by Richard Johnsonbaugh, Chicago State University.*

In L. W. Gates, *Summing square roots*, Math. Gaz., 64, No. 428, June 1980, pp. 86–89,

$$S_n = \frac{1}{24}\left[(8n+5)(4n+1)^{1/2} - 5\right]$$

is proposed as an approximation to

$$G_n = \sum_1^n t^{1/2}.$$

Prove that $\lim_{n\to\infty}(S_n - G_n)$ exists. What is the value of the limit?

E 2942. *Proposed by Vic Norton, Bowling Green State University.*

Show that

$$\det[a_{ij}]_1^n = \prod_{1\leqslant p<q\leqslant r}(x_q - x_p)^{n_p n_q},$$

where $n_1, n_2, \ldots, n_r$ are positive integers with sum n and

$$a_{ij} = \binom{i-1}{k-1} x_q^{i-k} \qquad (=0 \quad \text{if} \quad i<k)$$

with

$$q = \min\{p : j \leqslant n_1 + n_2 + \cdots + n_p\}, \qquad k = j - (n_1 + n_2 + \cdots + n_{q-1}.)$$

(*Note*: if each $n_q = 1$, this reduces to the standard Vandermonde formula.)

E 2943. *Proposed by Eric Anderson, student, St. Olaf College.*

For $n \geqslant 3$, let $Z_n = \{0, 1, 2, \ldots, n-1\}$,

$$A_n = \{(a,b,c) \mid a,b,c \in Z_n, a<b<c, a+b+c \equiv 0 (\text{mod } n)\},$$
$$B_n = \{(a,b,c) \mid a,b,c \in Z_n, a\leqslant b\leqslant c, a+b+c \equiv 0 (\text{mod } n)\}.$$

Let a_n and b_n be the number of elements of A_n and B_n respectively.

a. Show that $a_{n+3} = b_n$ and $b_n = a_n + n$.
b. Find an explicit formula for a_n in terms of n.

SOLUTIONS OF ELEMENTARY PROBLEMS

Watched Birds

E 2003 revised [1967, 720; 1968, 1009]. *Proposed by Robert Abilock, Brandeis University, Waltham, MA.*

N birds land at random in a field. Each bird watches its nearest neighbor. What is the expected number of unwatched birds?

No solution of this problem has appeared. In 1968, Michael Goldberg pointed out that the solution depends on the shape of the field and called attention to the corresponding one-dimensional problem, in which the birds land at random on a wire of finite length. (See J. Recreational Math., 1979, p. 227, problem 650 for a partial discussion.)

Solution of the one-dimensional case by the editors. The answer is $N/4$ if $N>3$. Label the birds sequentially from one end of the wire. Let D_i, $i=1,\ldots,N-1$, be the distance between the ith and $(i+1)$th bird. The key to the solution is to note that, while the D_i are not independently and identically distributed, they are exchangeable. The values of the D_i are distinct with probability 1 so we limit our discussion to that situation.

There are three kinds of birds: The 2 at the ends of the line, the 2 in the penultimate positions

and the remaining $N-4$. The bird at the left end of the line is unwatched exactly when $D_2 < D_1$, and because of exchangeability, $P(D_2 < D_1) = P(D_1 < D_2) = 1/2$. The next bird is watched, since the end bird watches it. For $3 \leqslant i \leqslant N-2$, bird i is unwatched exactly when $D_{i-2} < D_{i-1}$ and $D_{i+1} < D_i$. Because of exchangeability, this probability is $1/4$. The birds at the right end are like those at the left end. The probability is $1/N$ that a randomly chosen bird is in position i so we compute the probability that a randomly chosen bird is unwatched as $\frac{1}{2}(2/N) + 0(2/N) + \frac{1}{4}(N-4)/N = \frac{1}{4}$.

This result is independent of N so long as $N > 4$. If $N = 3$ there is only one penultimate bird and the required probability is $1/3$. For $N = 2$, there are no unwatched birds.

Since the number of birds watching is N, the number of doubly watched birds is the same as the number of unwatched birds, so the probability is $1/4$ that a bird is doubly watched.

Herbert (loc. cit.) has shown that, if $N > 3$, the probability is $2/3$ that a randomly chosen bird is watched by the bird that it is watching.

Recursively Defined Sequences

E 2835 [1980, 489]. *Proposed by Michael Golomb, Purdue University, Lafayette, IN.*

Let $-1 < a_0 < 1$, and define recursively $a_n = [\frac{1}{2}(1 + a_{n-1})]^{1/2}, n > 0$. Find the limits $(n \to \infty)$ of $A_n = 4^n(1 - a_n)$, $B_n = a_1 a_2 \cdots a_n$, $C_n = 4^n(B - a_1 a_2 \cdots a_n)$.

Solution by Gaston H. Gonnet, Department of Computer Science, University of Waterloo, Canada. Using the formula $\cos^2(x/2) = (1 + \cos x)/2$ and letting $\phi = \cos^{-1} a_0$, we find that $a_n = \cos(\phi/2^n)$ and consequently

$$A_n = 4^n(1 - a_n) = 4^n \left\{ \frac{\phi^2}{2 \cdot 4^n} - \frac{\phi^4}{24 \cdot 4^{2n}} + O(4^{-3n}) \right\} = \frac{\phi^2}{2} - \frac{\phi^4}{24^n} + O(4^{-2n}).$$

Thus $A_n \to \phi^2/2$.

For B_n, we find that by definition

$$\sin(\phi/2^n) B_n = \cos(\phi/2)\cos(\phi/4) \cdots \cos(\phi/2^n)\sin(\phi/2^n)$$

and telescoping with the use of $\sin(2x) = 2 \sin x \cos x$ we obtain

$$B_n = \frac{\sin \phi}{\sin(\phi/2^n)2^n} = \frac{\sin \phi}{\phi} + \frac{\sin \phi}{6 \cdot 4^n} + O(4^{-2n}).$$

Thus $B_n \to (\sin \phi)/\phi$, and $C_n = -(\phi \sin \phi)/6 + O(4^{-n})$, so $C_n \to -(\phi \sin \phi)/6$.

If $a_0 > 1$, then we may use cosh and sinh instead of cos and sin, and

$$\phi = \ln\left(a_0 + \sqrt{\left(a_0^2 - 1\right)}\right) = \cosh^{-1} a_0, \qquad A_n \to -\phi^2/2, \quad B_n \to (\sinh \phi)/\phi,$$

and $C_n \to (\phi \sinh \phi)/6$.

Also solved by M. Antony (Greece), David M. Bloom, Robert Breusch, Bela Brindza (Hungary), Paul Bruckman, Chico Problem Group, T. Fujinawa (Japan), John A. Gillespie, L. S. Hahn, Sidney Heller, W. T. M. Kars (the Netherlands), I. I. Kotlarski, Gesing Leung (Hong Kong), O. P. Lossers (the Netherlands), Nicholas A. Martin (Canada), L. E. Mattics, Jean Nordon (France), Otto G. Ruehr, T. Sekiguchi, A. Smuckler (Israel), Milan Šolc (Czechoslovakia), St. Olaf College Problems Group, Michael Vowe (Switzerland), and Michael Woltermann.

Six partially correct solutions were also received. J. Gillespie remarked that the limiting value of B_n was evidently known to the 16th century mathematician Viète (P. Beckmann, *A History of Pi*, The Golem Press, 1971). The editor notes that a similar problem appeared in Problem Proposal 2952, vol. 29 (1922) 81.

Ratio of Matrix Norms

E 2847 [1980, 671]. *Proposed by Emeric Deutsch, Polytechnic Institute of New York, Brooklyn.*

For the $n \times n$ real matrix $A = [a_{ij}]$, define

$$\|A\|_\infty = \max \sum_{j=1}^{n} |a_{ij}|\,;\ \|A\|^2 = \text{max proper value of } A^*A;\ |||A|||^2 = \sum_{i,j=1}^{n} |a_{ij}|^2.$$

Set $\operatorname{Re} A = \frac{1}{2}(A + A^*)$. Find

(i) $\max_{A\geq 0, A\neq 0}\|A\|_\infty/\|\operatorname{Re} A\|$; (ii) $\max_{A\geq 0, A\neq 0}\|A\|_\infty/|||\operatorname{Re} A|||$.

Here A^* denotes the transpose of A; $A \geq 0$ means A has no negative element.

Solution by Ron Adin, student, Haifa, Israel; Jorma Kaarlo Merikoski, University of Tampere, Finland; and Kam-chuen Ng & Yan-loi Wong, University of Hong Kong. The answer is (i) $2\sqrt{(n-1)}$ if $n \geq 2$ (and 1 if $n = 1$); (ii) $\sqrt{(2n-1)}$. To prove this, we need three lemmas.

(A) [**1**] *Let* $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$ *be n nonnegative numbers, not all zero. Then*

$$\max\left[\left(\Sigma x_i^2\right)^{1/2}/\Sigma x_i\right] = 1; \quad \max\left[\Sigma x_i/\left(\Sigma x_1^2\right)^{1/2}\right] = \sqrt{n}.$$

For the first maximum, it is necessary and sufficient that $x_1 = x_2 = \cdots = x_{n-1} = 0$. *For the second maximum, it is necessary and sufficient that* $x_1 = \cdots = x_n$.

(B) [**2**] *If* $A = (a_{ij})$, $B = (b_{ij})$ *are* $n \times n$ *matrices with* $0 \leq A \leq B$, *i.e.*, $\forall_{i,j}\ [0 \leq a_{ij} \leq b_{ij}]$, *then the dominant eigenvalue of B is at least as great as the dominant eigenvalue of A*: $0 \leq \lambda_{\max}(A) \leq \lambda_{\max}(B)$.

(C) *Let D be the symmetric* $n \times n$ *matrix* $(n \geq 2)$ *with zeros everywhere except in the first row and first column, and with first row* $[2d_1, d_2, \ldots, d_n]$. *The eigenvalues of D are the two numbers* $d_1 \pm (\Sigma d_i^2)^{1/2}$ *and (if* $n \geq 3$) 0.

The result (C) follows easily from the definition ($Ax = x\lambda$) of "eigenvalue."

Proof of (i) if $n \geq 2$. It is no restriction to assume that $\|A\|_\infty = \Sigma_j a_{1j}$. Denote $a_j = a_{1j}$; take A' as the matrix A, with the last $n-1$ rows replaced by zeros. Then $\operatorname{Re} A'$ has first row $[a_1, \frac{1}{2}a_2, \ldots, \frac{1}{2}a_n] = b$, first column b^* and zeros elsewhere. Moreover, $\operatorname{Re} A' \leq \operatorname{Re} A$. Now $\|A'\|_\infty = \|A\|_\infty$, and by (B), $\|\operatorname{Re} A'\| \leq \|\operatorname{Re} A\|$. Restricting attention, as we may, to matrices with zeros where A' has them, we see by (C) that our problem is to compute, for $a_1, a_2, \ldots, a_n \geq 0$ (and a_i not all 0), the maximum of $F/G = H$, where $F = \Sigma a_i$, $2G = a_1 + (\Sigma a_i^2)^{1/2}$. A short exercise in calculus ($\partial H/\partial a_1$) shows that H decreases with a_1, for each fixed $(n-1)$-tuple $a_2, \ldots, a_n$. Thus $\max H = \max[(a_2 + \cdots + a_n)/(a_2^2 + \cdots + a_n^2)^{1/2}] = \sqrt{(n-1)}$. □

Proof of (ii). We can again restrict attention to matrices that have zeros where A' has them. We find

$$\|A'\|_\infty/|||\operatorname{Re} A'||| = (\Sigma a_i)/|||\operatorname{Re} A'||| = \left[a_1 + \left(\frac{1}{2}a_2 + \frac{1}{2}a_2\right) + \cdots + \left(\frac{1}{2}a_n + \frac{1}{2}a_n\right)\right]/K,$$

$$K^2 = a_1^2 + \left(\frac{1}{2}a_2\right)^2 + \left(\frac{1}{2}a_2\right)^2 + \cdots + \left(\frac{1}{2}a_n\right)^2 + \left(\frac{1}{2}a_n\right)^2.$$

By (A), the ratio in question does not exceed $\sqrt{(2n-1)}$. Since this ratio is obviously attained by the n-tuple $[a_i] = [1, 2, \ldots, 2]$, the assertion follows. □

Merikoski noted the additional result

$$\left[|A|_p = \left(\Sigma_{i,k} a_{ik}^p\right)^{1/p}, A \geq 0\right]: \max_{A\geq 0, A\neq 0} \|A\|_\infty/|\operatorname{Re} A|_p = (2n-1)^{1-1/p}.$$

References

1. G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities, Cambridge University Press, 1952.
2. L. Mirsky, An Introduction to Linear Algebra, Clarendon, Oxford, 1955.

Series-Parallel Partially Ordered Sets

E 2853 [1981, 754]. *Proposed by John Barnden and David E. Daykin, Reading University, England.*

Consider the set C of all partially ordered sets (posets) constructible as follows. Start with the one-element poset. Then at each stage, in the obvious way, replace some element by one of the two two-element posets, using the inherited order relation with preference (if preference is needed). Show that the poset $P \in C$ if and only if every four-element subset of P (with inherited order) is in C.

Solution by A. G. Eins, Darmstadt, Germany. Consider the class N of *series-parallel* (s.-p.) posets (partially ordered sets) defined recursively as follows:

(i) every singleton is an s.-p. poset;

(ii) if P_1 and P_2 are s.-p. posets, then $P = P_1 \oplus P_2$ and $P' = P_1 + P_2$ are s.-p. posets.

[Here $P_1 \oplus P_2$ denotes the *ordinal sum* (each element of P_1 is smaller than every element of P_2) and $P_1 + P_2$ the *cardinal sum* (no element of P_1 is comparable with any element of P_2) of P_1 and P_2.] It is well known that N is exactly the class of finite posets not containing N as a subposet (see E. L. Lawler, "Graphical algorithms," Math. Centre Tracts, 81 (1976) 3-32). Thus the problem is solved if we can show that the classes N and C coincide. Now, by induction, it follows immediately from the definition that no member of C can contain N as a subposet.

Conversely, let $P = P_1 \oplus P_2$ be an s.-p. poset. By induction, assume $P_1, P_2 \in C$. To obtain P as a member of C, start with $p \circ \to \begin{smallmatrix} p_2 \\ p_1 \end{smallmatrix}$. Now transform p_i into P_i using the operations for C.

The case $P = P_1 + P_2$ is settled likewise starting with $p \circ \to p_1 \circ \quad \circ p_2$.

Also solved by the proposers.

$y^*(a) = \inf[\, y \mid \; y$ minimizes $f(a, y)]$

E 2854 [1980, 754]. *Proposed by James Chew, North Carolina Agricultural and Technical University, Greensboro.*

Let $f(x, y)$ be real valued and continuous on the unit square $0 \leqslant x, y \leqslant 1$. Set $y^*(a) = \inf\{y \mid y$ maximizes $f(a, y)\}$. Is $y^*(a)$ continuous? What if, for each $a, f(a, y)$ attains its maximum *uniquely* at $y^*(a)$?

Solution to part (i) *by the solvers listed*: No! Let $0 \leqslant z \leqslant 1$ and set $f(x, y) = (x - z)(y - z)$. Then $y^*(a) = 1$ if $a > z$, but $y^*(a) = 0$ if $a \leqslant z$.

Solution to part (ii) *by D. M. Bloom, Brooklyn College.* Yes! With the assumption that $y^*(a)$ is unique, then the graph of y^* is the intersection, over all $z, 0 \leqslant z \leqslant 1$, of the sets $S(z)$ defined by

$$S(z) = \{(a, y) \mid \; 0 \leqslant a \leqslant 1, \;\; 0 \leqslant y \leqslant 1, \;\; f(a, y) \geqslant f(a, z)\}.$$

Since f is continuous on a closed set, each set $S(z)$ is closed, the graph of y^* is closed and bounded. Thus y^* is continuous.

D. K. Mick asked for a reasonable characterization of the continuous functions $y^*(a)$ that arise in the context of E 2854.

Solved by Chico Problem Group, J. Cobb, J. M. Cohen & L. D. Day, C. D. Gerson, N. Glick, G. A. Heuer, D. Hill, E. Johnston, M. Josephy (Costa Rica), M. F. Kruelle, student, K. Lerasseur, student, G. S. Lessells (Ireland), B. Margolis (France), J. G. Mauldon, D. K. Mick, B. L. Montgomery, student, M. D. Meyerson, V. Pambuccian, student (Rumania), E. Posti (Finland), J. H. Riley, Jr., A. Shuchat, A. Smuckler (Israel), R. A. Struble, D. M. Wells, and the proposer.

Convergence of $a_n = a_{n-1}(a_{n-1} - 1)$

E 2859 [1980, 823]. *Proposed by Ulrich Faigle, Technische Hochschule, Darmstadt, Germany.*

Let the sequence $\{a_n\}$ of real numbers be defined by $a_n = a_{n-1}(a_{n-1} - 1)$ for $n \geq 2$. For what (initial) values of a_1 will this sequence converge?

Solution by Abraham Smuckler, Jerusalem College of Technology, Israel. If $-1 \leq a_1 \leq 2$, the sequence converges; otherwise, the sequence diverges monotonically to $+\infty$. The second assertion is obvious. If $a_1 = -1, 0, 1$, or 2, convergence is immediate. If $-1 < a_1 < 2$, the first assertion to be established is that there is an n for which $0 \leq a_n \leq 1$.

Proof. Case 1, $1 < a_1 < 2$. Then as long as a_n remains in this same interval, $1 < a_n < 2$, a_n is strictly decreasing ($a_{n+1} < a_n$). This shows that if $1 < a_1 < 2$, then a_n either approaches a limit, or else a_n eventually enters the interval $(0, 1]$. The first possibility is ruled out by the observation that the only numbers eligible to be limits of the sequence $\{a_n\}$ are the solutions $\{0, 2\}$ of $y = y^2 - y$. Case 2, $-1 < a_1 < 0$. In this case, $0 < a_2 < 2$; this case thus reduces to the case $0 < a_n \leq 1$; the meat of the issue resides in the case $0 < a_n < 1$.

If $0 < a_n < 1$, then $-\frac{1}{4} < a_{n+1} < 0$. The first observation to make in this case is that $|a_{n+1}| < a_n$. This is true because $(a_n - a_{n+1})/(2a_n) = 1 - \frac{1}{2}a_n < 1$. The second observation is that the recursion maps each of the intervals $(-\frac{1}{4}, 0)$, $(0, 1)$ ***into*** the other. The third observation is that if $0 < a_n < 1$, then $0 < a_{n+2} < a_n < 1$, since $a_{n+2}/a_n = 1 - 2a_n^2 + a_n^3 < 1$.

The final observation is that the odd-numbered members of the sequence move monotonically towards 0, and the even-numbered members do the same. Hence each subsequence approaches a limit, say x_0, y_0. These limits satisfy $x_0 = y_0^2 - y_0$, $y_0 = x_0^2 - x_0$; the conclusion is that $x_0 = y_0 = 0$.

There is a countable number of initial values a_1 for which convergence obtains in a finite (possibly large) number of steps, as noted by M. F. Kruelle. W. A. Smith noted that convergence is (can be) extremely slow.

The recursion $x_n = -x_{n-1}$ is a counterexample to the assertion of some solvers that an accumulation point of a recursion must be a fixed point. Indeed any initial value x_1 is an accumulation point of this counterexample.

Problem groups may enjoy the variations

$$a_{n+1} = (a_n - a)a_n; \quad z_{n+1} = z_n(z_n - 1)(z_n - 2).$$

Also solved by 55 other solvers, including the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by August 31, 1982. The solver's full post-office address should be on each sheet.

6383. *Proposed by Eliot Jacobson, University of Arizona, Tucson.*

Let $\phi(x)$ denote the totient function, and define recursively $\phi^r(x) = \phi(\phi^{r-1}(x))$. Let

$$A_m(n) = |\{x \leq n: \ \phi^k(x) = m, \text{some } k\}|.$$

Show that $\lim_{n\to\infty} A_m(n)/n$ exists for all $m > 0$, and calculate its value.

6384. *Proposed by Barry Powell, Kirkland, Washington.*

(a) For any positive integers m and n for which $m - n \neq 1$, prove that the set $\{m^p - n\}$, (where $p = 2, 3, 5, 7, \ldots$ is the set of primes) contains infinitely many composites.

(b)* Give an infinite set of n such that $(n + 1)^p - n$ is composite infinitely often ($p = 2, 3, 5, \ldots$).

6385. *Proposed by Mark D. Meyerson, U. S. Naval Academy, Annapolis, MD.*

Prove or disprove: Every simple closed curve in Euclidean space contains the vertices of a rectangle. (It is known to be true in the Euclidean plane.)

SOLUTIONS OF ADVANCED PROBLEMS

5413 [1966, 783; 1975, 85]. *Proposed by J. L. Selfridge and P. Erdős.*

Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers. Denote by $A_n^{(k)}$ the least common multiple of $a_n, a_{n+1}, \ldots, a_{n+k-1}$. Prove that the number of indices n for which $A_n^{(k)} < x$ is less than $cx^{1/k}$.

Comment by the editors. An article on this problem has appeared in Mat. Lapok, in Hungarian, by P. Erdős and E. Szemeredi. Let $1 \leqslant a_1 < a_2 \cdots$ be an infinite sequence A of integers; set $f(A, k, i) = \mathrm{LCM}[a_k, \ldots, a_{k+i-1}]$. If $F(A, x, i)$ is the number of integers k for which $f(A, x, i) < x$, then for $i = 2$, $F(A, x, i) < c_i x^{1/i}$, whereas for $i > 4$ this inequality is invalid.

Subsets of the Cantor Set

6213 [1978, 389]. *Proposed by C. G. Mendez, Metropolitan State College, Denver.*

Let G be an open dense subset of the Cantor set C. Is the boundary $\mathrm{Fr}(G)$ of G countable?

Solution by James P. Hatzenbuhler, Moorhead State University, and Michael Smith, Emory University, independently. The boundary of G may be uncountable. For each $i = 1, 2, 3, \ldots,$ let A_i be the discrete space $\{0, 1\}$. Then

$$C \simeq \prod_{i=1}^{\infty} A_i \simeq \left(\prod_{i=1}^{\infty} A_i\right) \times \left(\prod_{i=1}^{\infty} A_i\right) \simeq C \times C.$$

It thus suffices to find an open dense subset of $C \times C$ having uncountable boundary. Let $D = \{(a, b) \in C \times C \colon b \neq 0\}$. Then D is an open subset of $C \times C$, and since C is a perfect set, D is also dense in $C \times C$. The boundary of D is $C \times \{0\}$, and so the boundary is uncountable.

Also solved by F. S. Cater, Paul R. Chernoff, William J. Gorman Jr., Ronnie Levy, Mark D. Meyerson, John C. Morgan II, Jan Mycielski, Stephen Noltie, and the proposer.

Note. Many constructions were used, for example, with $I = [0, 1]$ and $a = \aleph_0$, the following. (1) $G = \{1, 2, 3\}^a \setminus \{1, 2\}^a$. (2) $G = C \setminus D$ where $C(D)$ is the base 4 numbers without 1 (without 1 and 2). (3) $C = 2^a$, G a dense open set of small Haar measure. (4) $G = g^{-1}[I \setminus C] \cap C, f$ the Cantor function. (5) $\mathrm{Fr}(G)$ a perfect set (homeomorphic with C). More generally, every uncountable compact metric space has a G with $\mathrm{Fr}(G)$ uncountable.

The Group $C_3 \times C_3 \times C_3$

6298 [1980, 408]. *Proposed by J. L. Brenner, Palo Alto, CA.*

If an arbitrary set of 19 lattice points (with integral coordinates) is given in euclidean 3-space, prove that some three have a centroid with integral coordinates. (This assertion is false if 19 is replaced by 18.)

Composite Solution by J. E. Cruthirds & L. E. Mattics, University of South Alabama, I. M. Isaacs, University of Wisconsin, and Forrest Quinn, undergraduate, California Institute of Technology. The proposal is equivalent to the following. In the group $G = C_3 \times C_3 \times C_3$ (direct product of three 3-groups), any set of ten elements contains three that add to 0. To see the equivalence of the two assertions, replace each coordinate in 3-space by its residue class mod 3. If the same element of G appears three times, the three preimages have an integral centroid. [The set $\{(0,0,0),$ $(0,1,0),$ $(0,0,1),$ $(1,0,2),$ $(1,2,0),$ $(1,2,2),$ $(2,2,2),$ $(2,0,2),$ $(2,2,0)\}$ of nine elements does not contain three that add to 0.]

Define $J^1 = \mathbb{Z}/(3)$, $J^n = J^{n-1} \times J^1$, so that $G \equiv J^3$. Choose a basis $v_1, v_2, \dots$ of J^n. With respect to this basis a *line* is any 3-set $\{k + av_i\}$, where k is some fixed point of J^n, i is fixed, and $a = 0, 1, 2$. A *plane* is any 9-set $\{k + av_i + bv_j\}$, $i \neq j$, where k is fixed, and $0 \leqslant a, b \leqslant 2$. A 3-space in J^n is similarly defined; if $n = 3$, any 3-space is all of J^3. A layer is a plane in J^3, all points of which have the same third coordinate. With respect to a fixed basis, any two lines are disjoint or identical; similarly for two planes.

We now broaden the definition of *line*, *plane*, etc., to mean any line with respect to any basis; any plane with respect to any basis. Then two "lines" can intersect in either one or no points. A counting argument shows that every point in $J^{(n)}$ lies on $\frac{1}{2}(3^n - 1)$ lines; every line lies on $\frac{1}{2}(3^{n-1} - 1)$ planes. The following facts are easily established. The three points of a line sum to $(0,0,0)$; any three points that sum to $(0,0,0)$ constitute a line. A line is determined by any two of its points. Two lines [planes] meet in 0 or 1 [3] points. The number of planes in J^3 that contain a fixed line L is four. (Each plane has six points outside L, and $4 \cdot 6 + 3 = 27$.)

The next result appears in the literature.

LEMMA 1. *Any set of* 5 *points in the same plane in* J^3 *contains a line.*

Proof. The easiest way to establish Lemma 1 is to construct all planar sets S that contain no line. If S has three or more points, say $S = \{k, k + v_1, k + v_2, \dots\}$, then $v_2 \neq 2v_1$ since no three points add to 0. Further, S does not contain any one of $k + 2v_1$, $k + 2v_2$, $k + 2v_1 + 2v_2$. Of the remaining three points $k + v_1 + v_2$, $k + 2v_1 + v_2$, $k + v_1 + 2v_2$ in the plane determined by the first three points of S, at most one can be included; since otherwise three points of S add to 0. □

COROLLARY. *If* $|T| = 10$, *and if* T *contains no line, then every layer contains at least two points of* T (since $4 + 4 + 1 = 9 < 10$).

Now suppose T contains 10 points of J^3, and suppose T contains no line. Then every 3 points of T determine (and lie in) a plane. Let v, w be distinct points of T; let L be their line. There are precisely four planes through L. All 10 points of T must lie in one or another of these planes. By Lemma 1, none of these planes contains more than 4 points of T; so each of the planes contains exactly 4 points of T (since $|\{v, w\}| + \frac{1}{4}[|T| - |\{v, w\}|] = 4$). In other words, every plane that contains as many as 2 points of T contains 4 points of T. □

Quinn found the following extension to J^4. Any set of 21 elements in J^4 must contain a line. As a corollary, any collection of 41 lattice points in 4-dimensional space must contain some three that are collinear. This assertion is not true, if 41 is replaced by 40.

H. Harborth (W. Germany) called attention to his article, "Gitterpunktprobleme," which appeared in J. Reine Angew. Math, 309 (1979) 49–155; it discusses similar questions. In particular, the assertion 6298 is anticipated.

Also solved by Loren C. Larson, To Wai-Ping (Hong Kong), and the proposer.

Proposal 6298 lies in the field "additive group theory," which is a generalization of some portions of number theory. The first paper in this new field was published by P. Erdös, A. Ginzburg, and A. Ziv [Bull. Res. Council Israel, 10F (1961) 41–43]. See the survey paper of H. B. Mann, Bull. Amer. Math. Soc., 79 (1973) 1069–1075. For Lemma 1, see H. B. Mann, *Two addition theorems*, J. Combin. Theory, 3 (1967) 233–235.

A weaker form of the statement 6298 was suggested by Rumania for use in the 1977 International Olympiad, but was not used.

MISCELLANEA

71. Mathematics is the science of saving thought.

—G. A. Miller, this MONTHLY, 15 (1908) 197.

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Mathematics of Cell Electrophysiology. By Jane Cronin. Lec. Notes in Pure and Applied Math., vol. 63, Marcel Dekker, New York, 1981. pp. 144, $19.75. ISBN 0-8247-1157-2.

STUART HASTINGS
Department of Mathematics, SUNY at Buffalo, Buffalo, NY 14214

Biology, despite the efforts of many researchers in the past decade, remains on the frontier of the areas to which mathematics has been applied successfully. Only in genetics is there a large amount of significant mathematical literature, due largely to pioneering work earlier in the century by J. B. S. Haldane, R. A. Fisher, and S. Wright. Extensive modeling has been done in related fields, such as ecology and population biology, but this work is difficult to check quantitatively and has not yet gained wide acceptance with biologists.

In more experimental subjects, such as physiology, techniques of measurement are inadequate to support quantitative predictions obtained from mathematical models which are based on first principles. No counterpart to Newton's inverse square law or the Navier-Stokes equations has emerged to explain biological organization of physiological processes.

There are, however, a few areas in physiology where experimental results have become precise enough to generate quantitatively accurate models with significant capacity for prediction. One is biological fluid mechanics, as used for example in studies of blood flow. In this instance classical techniques developed independently of biology have proved adaptable to biological contexts. An excellent introduction can be found in [**9**].

Another area (the only other?) where mathematical approaches have clearly been successful is the subject of the book under review, neurophysiology. Here the applications of mathematics are wide-ranging and include efforts to understand the macroscopic organization of the nervous system [**12**] as well as studies at the single cell and subcellular levels. Multicellular theories are still in an early stage of development, but in considering electrical activity in a single cell we come as close as anywhere in physiology to a mathematical model of fundamental importance, the celebrated partial differential equations of A. L. Hodgkin and A. F. Huxley describing the propagation of an electrical impulse along a nerve axon [**6**].

These equations were developed about thirty years ago on the basis of extensive and ingenious experimental measurements. The importance of this work was quickly recognized by biologists; Hodgkin and Huxley were awarded the 1963 Nobel prize in physiology and medicine. However, not until around 1970 did mathematicians begin to explore the many problems in differential equations which spring from this research. Surveys of some of this mathematics can be found in [**5**], [**7**], [**10**], and [**11**], as well as in the somewhat more specialized volume of Cronin.

It must be emphasized that the significance of the Hodgkin-Huxley equations in nerve studies is not as great as that of the Navier-Stokes or Newtonian models in their realms of applicability. The precise form of the Hodgkin-Huxley equations was derived using considerable empirical curve fitting, because there was no adequate theory of the physiology of a nerve cell membrane. This major gap in our understanding of the nervous system still exists, and, as a result, alternatives to the Hodgkin-Huxley formulation continue to appear. The importance of their model is the extent to which it allows detailed predictions of complicated physiological phenomena. For example, in the original paper presenting the model, it was shown how the speed of propagation of a nerve impulse can be obtained from measurements of much more elementary processes. (This basic point seems to have been misunderstood by C. Zeeman, whose model of the nerve is rightly criticized by Cronin.) A recent striking example of a prediction obtained from the Hodgkin-Huxley equations appears in [**3**]. This is an instance where behavior which was first suspected because of numerical analysis and analytical studies of the equations [**4**], [**8**] was later found experimen-

tally. Surely this is one of the major goals of a mathematical biologist.

The book by Cronin begins with a clear discussion of the physiological background of the Hodgkin-Huxley equations. The second chapter is devoted to related models of electrical activity in the heart, due to D. Nobel and coworkers. Perhaps this research is not sufficiently well established to warrant the attention given it by Cronin, but it serves to illustrate her point that quantitatively realistic descriptions of the nervous system may have to be extremely complicated. The response of many mathematicians to this complexity has been to emphasize simpler qualitative models, in an effort to develop principles which may then be checked, at least numerically, on more realistic sets of equations.

The Hodgkin-Huxley model is a coupled nonlinear system of four differential equations which, in its most general form, contains both derivatives with respect to time and derivatives with respect to space. There are special solutions which are independent of the spatial variable and hence satisfy ordinary differential equations, since only the time derivatives are nonzero. It is possible to achieve spatial homogeneity in a nerve axon artificially in a laboratory, and experiments done in this setting were used extensively by Hodgkin and Huxley. However, in order to describe signal propagation along the axon it is necessary to allow for spatial variation of the state variables.

Much of the book under review is devoted to problems about periodic solutions of ordinary differential equations. Physically this subject is more important in relation to cells where spatial differences are insignificant than to the preparations on which Hodgkin and Huxley did their experiments. However, many of the same principles apply. The existence and stability of periodic solutions for complicated systems of ordinary differential equations is a difficult problem, and in the longest of the five chapters a theory is presented which the author hopes will contribute to dealing with these questions. One may ask, however, whether the considerable diversion into dynamical systems theory is justified in the absence of any conclusive applications of this theory.

Moreover, the single-minded attention to spatially homogeneous solutions may mislead those unacquainted with the field. This is not to say that these problems are uninteresting or unimportant. Indeed, the discovery of new physiological behavior which I mentioned above came about through numerical and theoretical bifurcation studies of just the kind of system Cronin discusses. I believe, however, that most researchers today are interested in the Hodgkin-Huxley equations as partial differential equations, one of a class of so-called reaction-diffusion systems which are being widely examined. These are systems of p.d.e.'s in which diffusion processes interact with nonlinear dynamics.

By the dynamics of a system I mean the terms which determine the behavior of the spatially homogeneous solutions. In many cases these terms represent chemical reactions. As noted above, such solutions satisfy a system of ordinary differential equations, and one can ask, for example, if this system has periodic solutions. (For the Hodgkin-Huxley equations in their original form the answer seems to be "no.") Other properties are also important, such as the existence of some sort of threshold in the response of the nerve to incoming stimuli. One of the important challenges in this field is to relate dynamical properties to the solutions of the full system of p.d.e.'s.

Among topics of current interest in relation to the Hodgkin-Huxley equations, stability is probably the most important. Roughly speaking, a solution is stable if small changes in the initial conditions do not significantly change the long-term behavior of the solution. Unstable solutions probably cannot be observed and are generally uninteresting physiologically. It has never been proved that the Hodgkin-Huxley equations have stable nonconstant solutions.

An important development in recent years has been the observation that other physical systems behave like nerve cells. Perhaps the most carefully studied example is a chemical reagent which displays oscillatory behavior under some conditions, and in other configurations supports wave propagation in a manner very similar to the nerve axon. This reagent is usually associated with the name of Zhabotinsky, and some of its important properties were discovered by A. Winfree [**13**]. A mathematical model of the Zhabotinsky reagent due to R. Field and R. Noyes [**1**] has been found

to have many properties in common with the Hodgkin-Huxley equations, even though the nonlinearities in the two systems look very different. Many interesting phenomena, such as two-dimensional rotating spiral patterns, have been seen in experiments on the Zhabotinsky reagent, prompting theoretical and experimental searches for similar behavior in the nervous system [**12**]. The mathematical theory behind this kind of behavior is incomplete, and this is one reason why the problem of characterizing spatial properties of solutions in terms of dynamics is important.

Readers of this small volume will be introduced to the physiological background of the Hodgkin-Huxley equations and related systems. They will be exposed to the complexity which seems to be a necessary component of quantitatively accurate nerve models, and perhaps appreciate the reasons for considering simplified models for most analysis. However, it should be kept in mind that the study of nerve equations goes beyond a consideration of ordinary differential equations, and, when attention is broadened to include spatial effects, a much richer variety of solutions is opened for investigation.

References

1. R. J. Field and R. M. Noyes, Limit cycle oscillations in a model of a real chemical reaction, J. Chem. Phys., 60 (1974) 1877–1884.

2. P. C. Fife, Mathematical aspects of reacting and diffusing systems, Lecture Notes in Biomathematics, vol. 28, Springer-Verlag, New York, 1979.

3. R. Guttman, S. Lewis, and J. Rinzel, Control of repetitive firing in squid axon membrane as a model for a neuron oscillator, J. Physiol., 1980.

4. B. D. Hassard, Bifurcation of periodic solutions of the Hodgkin-Huxley model for the squid giant axon, J. Theoret. Biol., 71 (1978) 401–420.

5. S. P. Hastings, Some mathematical problems from neurobiology, this MONTHLY, 82 (1975) 881–895.

6. A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol., 117 (1952) 500–544.

7. J. Rinzel, On repetitive activity in nerve, Federation Proceedings, 1979.

8. J. Rinzel and J. N. Miller, Numerical calculation of stable and unstable periodic solutions to the Hodgkin-Huxley equations, Math. Biosci., 1980.

9. S. I. Rubinow, Introduction to Mathematical Biology, John Wiley, New York, 1975.

10. A. C. Scott, Neurophysics, Wiley-Interscience, New York, 1977.

11. W. C. Troy, Mathematical modeling of excitable media in neurobiology and chemistry, in Theoretical Chemistry, vol. 4, Academic Press, New York, 1978.

12. H. Tuckwell and R. Miura, A mathematical model for spreading cortical depression, Biophys. J., 23 (1978) 257–276.

13. A. T. Winfree, Stably rotating patterns of reaction and diffusion, in Theoretical Chemistry, vol. 4, Academic Press, New York, 1978.

MISCELLANEA

72.

'Cos some of us don't know the way
To say the name of André Weil,
It leaves the rest of us the while
Confusing him with Hermann Weyl.

RICHARD K. GUY

The Mathematical Association of America

CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Academic Press, Inc.
Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
Harcourt Brace Jovanovich, Inc.
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.

MAA PLACEMENT TEST PROGRAM

EVERY STUDENT BELONGS

Let the MAA Placement Test Program help you match entering students with beginning mathematics courses according to ***training*** and ***ability,*** rather than transcripts and credentials. PTP tests are constructed by panels representing a broad spectrum of institutions and are carefully pretested. Information about pretesting scores and placement experience of a variety of participating institutions is published periodically in the PTP Newsletter.

Your institution can have unlimited use of annually updated MAA Placement Tests on

- **Basic Mathematical Skills**
- **Basic Algebra**
- **Advanced Algebra**
- **Trigonometry/Elementary Functions**
- **Calculus Readiness**

and also a subscription to the PTP Newsletter for one modest annual subscription fee.

For information write to:
The Mathematical Association of America
Department PTP
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE BICENTENNIAL TRIBUTE TO AMERICAN MATHEMATICS

Edited by DALTON TARWATER

This volume is based on the papers presented at the Bicentennial Program of the Association on January 24–26, 1976. In addition to the major historical addresses, the papers cover the following panel discussions: Two-Year College Mathematics in 1976; Mathematics in Our Culture; The Teaching of Mathematics in College; A 1976 Perspective for the Future; The Role of Applications in the Teaching of Undergraduate Mathematics.

The following is a list of the Panelists and the Authors: Donald J. Albers, Garrett Birkhoff, J. H. Ewing, Judith V. Grabiner, W. H. Gustafson, P. R. Halmos, R. W. Hamming, I. N. Herstein, Peter J. Hilton, Morris Kline, R. D. Larsson, Peter D. Lax, Peter A. Lindstrom, R. H. McDowell, S. H. Moolgavkar, Shelba Jean Morman, C. V. Newsom, Mina S. Rees, Fred S. Roberts, R. A. Rosenbaum, S. K. Stein, Dirk J. Struik, Dalton Tarwater, W. H. Wheeler, A. B. Willcox, W. P. Ziemer.

Individual members of the Association may purchase one copy of the book for $10.00; additional copies and copies for nonmembers are priced at $15.00 each. (Orders for under $10.00 must be accompanied by payment. Prepaid orders will be delivered postage and handling free.)

Orders should be sent to:

MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W. Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 5 May 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

EDITOR

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

WHAT IS AN ANSWER?

HERBERT S. WILF
Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104

In many branches of pure mathematics it can be surprisingly hard to recognize when a question has, in fact, been answered. A clearcut proof of a theorem or the discovery of a counterexample leaves no doubt in the reader's mind that a solution has been found. But when an "explicit solution" to a problem is given, it may happen that more work is needed to evaluate that "solution," in a particular case, than exhaustively to examine all of the possibilities directly from the original formulation of the problem. In such a situation, other things being equal, we may justifiably question whether the problem has in fact been solved.

Examples of this sort can turn up anywhere, but here we will concentrate on problems in combinatorial mathematics, specifically those of the type "how many—are there?" Such enumeration problems lie at the heart of the subject, and it is important to be able to recognize solutions when they appear. The point, of course, is that sometimes the "answer" is presented as a formula that is so messy and long, and so full of factorials and sign alternations and whatnot, that we may feel that the disease was preferable to the cure.

An answer to such an enumeration question may be given by means of a generating function, a recurrence relation, or by an explicit formula. Each of these is, in essence, just an algorithm for the computation of the counting sequence that is to be determined.

How do we judge the usefulness of such answers? Obviously we might be able to do many things with the answer, such as to make asymptotic estimates, to discover congruence relations, to delight in its elegance, and so forth. We're going to restrict attention here to the appraisal of solutions from the point of view of how easily they allow us to calculate the number of objects in the set that is being studied.

The quality of such formulas should therefore be judged by the usual combination of esthetic and quantitative benchmarks that are used on algorithms. In particular, the quantitative criterion is the computational complexity: the amount of work required to get an answer. We suggest here that the same criterion should be applied to enumeration formulas. We will see that a corollary of this attitude is that our decision as to what constitutes an answer may be time-dependent: as faster algorithms for listing the objects become available, a proposed formula for counting the objects will have to be comparably faster to evaluate.

For concreteness, suppose that for each integer $n > 0$ there is a set S_n that we want to count. Let $f(n) = |S_n|$ (the cardinality of S_n), for each n.

Suppose further that a certain formula has been found, say

$$f(n) = \text{Formula}(n) \qquad (n = 1, 2, \ldots) \tag{1}$$

in which Formula(n) may involve various multiple summation signs extending over various sets and various complicated summands, etc.

In order to evaluate the "answer" (1), let's look at the competition. To insure my own immortality in the subject, I am now going to show you a simple formula that "answers" all such questions at once. If you're ready, then, here it is:

$$f(n) = \sum_{S_n} 1. \tag{2}$$

Well, anyway, the summand is elegant, even if the range of summation is a bit untidy.

Now (2) is unacceptable in polite society as an answer, despite its elegant appearance, because it is just a restatement of the question, and it does not give us a tool for calculating $f(n)$ that we

In addition to his other accomplishments (see this volume, p. 4) the author is currently one of the associate editors of this MONTHLY.

didn't have before. It does illustrate an important point, though: there is no counting problem for which a formula does not already exist, namely (2).

A first criterion for evaluating an "answer," then, might be that "Formula(n)" should be an improvement over the insightful contribution (2).

What is an improvement? As noted previously, we are considering a formula as an invitation to compute numbers; as an algorithm, if you will, that describes a sequence of steps that will lead to an answer, and it will be useful if less effort is required to use it than to use some other formula or algorithm.

What is effort? Now we're approaching firm ground. The theory of computational complexity has been developed rapidly in recent years and there is no shortage of rigorous standards for measuring computing effort. Usually we adopt some indivisible units of labor, such as multiplications or divisions of numbers in a certain size range, or bit operations, or function evaluations, etc., and then we express the complexity of the computation by counting how many units of each kind of labor need to be performed in order to get the job done.

How do we compare the complexity of evaluating Formula(n) with (2)? Equation (2) calls for the straightforward listing algorithm: produce all of the members of the set S_n and count them. Hence the complexity of (2) is equal to the amount of computational effort that is required to produce all members of S_n.

Next, recall that we need a little more generality in the format of the "answer." Formulas are not always the way solutions are given. Two other useful methods, for example, are the method of generating functions, and the use of recurrence relations. So, in all cases, we will consider the computational complexity of using whatever tool is given for the purpose of computing $f(n)$, whether it be a formula, a recurrence relation, or other.

The functions that will be compared are, therefore:

Count(n) = the complexity of the algorithm for calculating $f(n)$, whether it be given by a formula, an algorithm, et cetera, and

List(n) = the complexity of producing all of the members of the set S_n, one at a time, by the speediest known method, and counting them.

DEFINITION 1: We will say that a solution of a counting problem is *effective* if

$$\lim_{n\to\infty} \frac{\text{Count}(n)}{\text{List}(n)} = 0.$$

What we are saying is that a formula or whatever is an effective solution of a problem if the effort involved in counting the members of S_n by means of that formula is asymptotically small compared to the effort of constructing all of the members of S_n, by the best-known algorithm, and counting them.

EXAMPLE 1: In the theory of numbers it is often said (see [**5**, p. 344]) that there is "no formula for $\pi(n)$ (the number of primes less than or equal to n)." Now, of course, there is such a formula, namely the one given by (2) above. In the light of the definition, though, we can give the precise meaning of the above statement in the form "there is no effective solution to the question of enumerating the primes less than or equal to n."

How good would a formula or other algorithm for $\pi(n)$ have to be in order to qualify as effective? To list all of the primes between 1 and n we can use one of the fast forms of the sieve of Eratosthenes, such as [**4**]. That algorithm operates in time $Kn/\log\log n$. Hence a "formula" for $\pi(n)$ would have to be faster than that in order to satisfy the condition of definition 1.

Again, in complexity theory one speaks of computational problems as being "easy" or "hard" depending on whether the complexity of finding a solution is or is not of polynomial growth in the length of the input bit string. This distinction was first made by Jack Edmonds, and it is responsible for much of the explosive growth in complexity theory today.

To take a leaf from that book, then, a formula "solves" an enumeration problem if the formula yields numbers after "easy" calculations. To make this assumption meaningful we will now restrict attention to those problems in which $f(n)$ grows faster than any polynomial in n. We will

call this the class of $\nu\pi$ (νot πolynomial) problems.

More precisely, then, we propose

DEFINITION 2: We will say that a problem in the class $\nu\pi$ has been p-solved if Count(n) is of polynomial growth as $n \to \infty$.

Respectively, then, we may speak of an enumerative problem as being "unsolved," "effectively solved," or "p-solved."

It is interesting to note that the only input to the computation of $f(n)$ is n itself, and so the length of the input bit string is about log n bits. To ask for polynomial growth in n is therefore to allow exponentially rapid growth according to the customary standards of complexity theory. Nonetheless, this weaker criterion seems quite stern enough, as we will see from the examples.

EXAMPLE 2: For quite a clearcut situation, consider the number of permutations of n letters that have no fixed points. The well-known, and very elegant, solution is that $f(n)$ is equal to the nearest integer to $n!/e$, or equivalently, that

$$f(n) = \sum_{j=0}^{n} (-1)^j \frac{n!}{j!} \qquad (n = 1,2,\ldots). \tag{3}$$

This shows first of all that we are indeed dealing with a problem in the class $\nu\pi$, and second, that this formula really is an answer because we can compute from it in polynomial time, i.e., the problem is p-solved.

EXAMPLE 3: Next, here's an example from graph theory. If we ask for the number of unlabelled graphs on n vertices, we quickly find ourselves using the Frobenius-Burnside theory of group action on a set (e.g., [**1**]). This theory counts the equivalence classes of a set that is acted on by a group of permutations, in terms of the sets of elements that are fixed by the permutations in the group.

Obviously, if a certain permutation has $s(i)$ cycles of length i, then $s(1) + 2s(2) + 3s(3) + \cdots = n$ is a partition of the integer n. The answers that emerge from the Frobenius-Burnside lemma (or from the more general theory of Pólya [**2**]) depend only on the cycles, and so they depend only on the partitions of n. Typically, then, an answer that comes from this theory involves a sum of a more-or-less elementary function extended over all partitions of n.

For instance, the number of nonisomorphic unlabelled graphs on n vertices is exactly ([**2**])

$$f(n) = \frac{1}{n!} \sum_{\pi} \left(\frac{n!}{1^{s(1)}s(1)!2^{s(2)}s(2)!\cdots} \right) 2^{g(\pi)} \tag{4}$$

where

$$g(\pi) = (1/2) \left\{ \sum_{i,j=1}^{n} \gcd(i,j)s(i)s(j) - \sum_{k} s(2k+1) \right\}$$

and π: $n = s(1) + 2s(2) + 3s(3) + \cdots$ runs through the partitions of n.

Evaluation of g, for a given partition of n, is clearly of polynomial complexity. The number of terms in the sum (4) is $p(n)$, the number of partitions of n, and this is well known to grow like $A\exp(B\sqrt{n})/n$. Thus the evaluation of the formula (4) is not a polynomial-time job, since Count(n) grows faster than $A\exp(K\sqrt{n})$ for some constant K.

How big is $f(n)$ itself? The asymptotic behavior of the number of graphs on n vertices is

$$f(n) \sim 2^{\binom{n}{2}}/n! \qquad (n \to \infty).$$

Hence the criterion of Definition 1 is amply satisfied: the answer given by Frobenius-Burnside's lemma is a drastic improvement over (2). This problem is therefore effectively solved. It is not p-solved, however, in the sense of definition 2, because the formula (4) calls for more than a polynomial amount of work.

This raises the interesting question of whether there exists a polynomial-time formula for this

problem. Indeed, can one describe a reasonable and natural family of combinatorial enumeration problems for which there is provably no polynomial-in-n time formula or algorithm to compute $f(n)$? Further, is there any relationship between the intractability of the isomorphism problem for unlabelled graphs and the apparent un-polynomial-ness of the counting problem? These questions are reminiscent of, but not identical to, the concept of $\#P$-completeness discussed in [3].

In addition to the theoretical interest of the questions of computational complexity of formulas, there are practical algorithmic consequences too. Frequently we find that in order to carry out a combinatorial algorithm we need some values of the relevant counting functions. In [6], for example, there is an algorithm that needs to know the number of unlabelled graphs on n vertices before it can begin to do its job. Other algorithms need Bell numbers, and so forth.

EXAMPLE 4: We'll conclude with the example of the partition function itself. Since the sum (4) extended over partitions was hard, what about the sum of 1 over the partitions of n?

That one is easy, because we have any number of simple recurrence relations, generating functions, and so forth, from which the values of $p(n)$ can be rapidly calculated. For instance

$$np(n) = \sum_{k=1}^{n} p(n-k)\sigma(k) \qquad (n \geqslant 1, p(0) = 1) \tag{5}$$

where $\sigma(k)$ is the sum of the divisors of k. This raises at least the hope that the previous problem may have a polynomial time formula also, and it raises the question of describing the class of functions f of partitions that have the property that the function $g(n)$, obtained by summing f over all partitions of n, can be evaluated in polynomial time.

Added in proof: After reading a preprint of this article, Jeffrey Lagarias and Andrew Odlyzko of Bell Laboratories found an algorithm that computes $\pi(x)$ in time $O(x^{\frac{3}{5}+\epsilon})$: a true "formula" for $\pi(x)$.

References

1. N. G. de Bruijn, Pólya's theory of counting, Chapter 5 in Applied Combinatorial Mathematics, E. F. Beckenbach, editor, Wiley, New York, 1964.
2. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta Math., 68 (1937) 145–254.
3. Michael R. Garey and David S. Johnson, Computers and Intractability, a Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
4. Paul Pritchard, A sublinear additive sieve for finding prime numbers, Comm. ACM, 24 (1) 1981, pp. 18–23.
5. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 3rd ed., Oxford, the Clarendon Press, 1954.
6. John D. Dixon and Herbert S. Wilf, On the random selection of unlabelled graphs, in preparation.

A MATHEMATICAL ANALYSIS OF THE GAME OF JAI ALAI

LOUISE E. MOSER
Department of Mathematics, California State University, Hayward, CA 94542

1. Introduction. Jai Alai is a ball game which evolved during the seventeenth century in the Basque provinces of Spain. It is played in a court with three walls by eight players (singles) or eight teams of two players (doubles). The ball, called the pelota, is thrown against the front playing wall with a curved reed basket, called the cesta. It can bounce off the side wall, the back

---

Louise Moser received her Ph.D. from the University of Wisconsin in 1970. Her main area of research and publication has been 3-dimensional topology and knot theory. She is currently a professor at California State University, Hayward, where she teaches courses in Mathematics and Computer Science.

wall and/or the floor within certain specified areas. When a player fails to return the ball to the front wall before it strikes the floor a second time, a point is scored by his opponent. The game is played in round robin fashion with the winner of a point continuing to play against the next player until he loses, at which time he takes his place at the end of the bench and waits for his turn to come up again.

Like horse racing, Jai Alai operates under a pari-mutuel system of betting. A wager to win is a bet that a certain player or team will finish first. A wager to place is a bet that a certain player or team will finish at least second. A wager to show is a bet that a certain player or team will finish at least third. The exacta or perfecta bet is a bet on two players or teams, one to finish first and the other to finish second, in precisely that order. The quiniela bet, which originated in Jai Alai, is a bet on two players or teams to finish first or second, in either order.

The study presented in this paper was motivated by the need for Jai Alai gamblers to know the advantages and disadvantages of a player's starting in various post positions because, according to the way the game is set up (see the rules described below), a player's chances of winning depend on his starting post position. The investigation of Jai Alai is, in itself, a useful example for introducing the concepts of a game tree and its probabilistic analysis to students since (1) it is a real world example whose results can be seen to be of importance and (2) the analysis is not trivial and its results are not obvious, and yet a complete analysis is possible. Furthermore, it is an especially nice application of the programming language Pascal. It is just the type of problem for which Pascal was designed and for which it is awkward to use Fortran.

2. Rules of the Game. In games with eight post positions, the seven point system of scoring is commonly used. According to the rules of this system [2] one point is awarded for each point scored through the first complete round of play, that is, in the first seven matches. After each player has played once, two points are awarded for each point scored. Game point is one point less than the number of players or teams, that is, seven points. The first player or team to reach seven or more points is declared the winner. Place and show are awarded to the players or teams with the second and third highest numbers of points, respectively. In case of ties there is a playoff. The playoff rules for place and/or show vary according to the number of players or teams tied as described below.

- Two-way tie.
  - For place and show. Place is awarded to the winner of the first point played, and the loser takes show.
  - For show only. Show is awarded to the winner of the first point played.
- Three-way tie.
  - For place and show. Place is awarded to the winner of two points. The loser of the first point played is eliminated from further play for place. Show is awarded to the winner of one point.
  - For show only. The winner of two points takes show. The loser of the first point played is eliminated from further play.
- Four-way or six-way tie.
  - The playoff is decided through elimination. The first two participants play for the first point, the next two participants play for one point and, in the case of a six-way tie, the remaining participants play for one point. Losers are eliminated, and winners play for the additional point or points, under the rules governing two-way or three-way ties, as the case may be.
- Five-way or seven-way tie.
  - If the players or teams are tied without a point or with one point, the playoff is for one point less than the number of players or teams tied, that is, for four or six points. If they are tied at two or more points, the playoff is for the full game number of points. The scores of the remaining players continue from the number of points at which they were tied when a winner was declared.

There are, however, two overriding rules which apply at all times during playoffs.

• Playoffs follow the players' or teams' order on the bench (not their post positions) at the time a winner is declared.

• If at any time during a playoff a player or team reaches the full number of points for a game, the player or team is immediately awarded place or show as the case may be, and the remaining players or teams are eliminated from further play for that position.

3. Analysis by Construction of the Game Tree. To avoid the biases of random sampling, an exhaustive computer analysis of the game (rather than a statistical examination of the records) was done. All of the possible 422,384 games in the game tree were considered.

We begin with a generalization of the game in which there are n players, one point is awarded for each point scored in the first $n - 1$ matches and two points are awarded for each point scored thereafter, and the first player to score a total of $n - 1$ points wins the game.

As a model of the game and our program we construct a directed binary tree (the win tree) which represents all the different ways in which the various players can win a game. The root of the tree corresponds to the initial match between the players in post positions one and two. Each node of the tree is labeled with the winner of a match between two consecutive players or teams. The edges are directed and labeled with probabilities. Each of the possible game outcomes can be described by a sequence of winning post positions at the nodes on a path from the root of the tree to a particular end node (or leaf) of the tree. The leaves represent all possible ways of winning.

If s is the number of leaves and t is the number of nodes, then $t = 2s - 1$. To prove this we let r denote the number of internal nodes and e the number of edges. Then $t = r + s$. Since each node except the root lies at the bottom of exactly one edge, $e = t - 1$. Furthermore, since each internal node is at the top of exactly two edges, $e = 2r$. Hence, $2r = t - 1 = r + s - 1$ from which it follows that $r = s - 1$ and finally $t = 2s - 1$. For a four player game one easily sees that the number of leaves of the win tree is 22, and so the number of nodes is 43. For the standard eight player game the number of leaves of the win tree was found by a computer to be 422,384. Thus, the number of nodes in this case is 844,767.

In our algorithm we visit these nodes in the preorder traversal, that is, first we visit the root of the tree or subtree, next we traverse the left subtree in preorder and then we traverse the right subtree in preorder. The height of a node, that is, the number of edges in a path from the root to the node, corresponds to the number of matches up to that point and determines whether one or two points are awarded for each point scored. The height of the tree, that is, the maximum of the heights of all the leaves, represents the maximum size of a stack used in our algorithm which traverses the tree.

The height of the win tree is given by the formula $n(n - 1)/2$. To prove this we show first that the height is no more than this number. Since one point is awarded in the first $n - 1$ matches and two points are awarded in each of the remaining matches, the total number of points accumulated by all the players at the end of the game is odd if n is even and is even if n is odd. If there were n players tied at $n - 2$ points each before the final match, then the total number of points of all the players would be $n(n - 2) + 2$ which is even if n is even and odd if n is odd. This contradiction shows that such a situation cannot occur.

Thus, the maximum possible number of matches is no more than the number which would occur if there were $n - 1$ players tied at $n - 2$ points each and one player with $n - 3$ points before the final match. The total number of points of all the players after the final match then is

$$(n - 1)(n - 2) + (n - 3) + 2 = n^2 - 2n + 1,$$

and there are

$$((n^2 - 2n + 1) - (n - 1))/2 = (n^2 - 3n + 2)/2$$

matches in which two points are awarded per match. Hence, the maximum possible number of

matches is

$$(n-1)+(n^2-3n+2)/2=n(n-1)/2.$$

To see that this maximum actually occurs we must again consider the cases n even and odd. If n is even ($n=2k$ for some positive integer k), then there are $n(n-1)/2$ matches in the game with the following sequence of winning post positions $1\ldots 1n\ldots nk\ldots k(n-1)\ldots(n-1)(k-1)\ldots(k-1)\ldots 2\ldots 2(k+1)\ldots(k+1)1$. In this sequence there is an initial group of $n-2$ ones followed by $n-1$ groups with $(n-2)/2$ elements each and the final one. If n is odd, then there are $n(n-1)/2$ matches in the game with the sequence of winning post positions $1345\ldots n21345\ldots n2\ldots 1345\ldots n2$. In this case there are $(n-1)/2$ groups with n elements each. We have then that the height of the win tree is $n(n-1)/2$. In particular, for a game with four players the height of the win tree is 6, and for an eight player game the height of the win tree is 28.

Furthermore, the height of the extended tree is $n(n-1)/2+2=(n^2-n+4)/2$. This follows because of the rule that a playoff terminates if a player or team reaches the full number of points for a game. To see this note that the maximum possible number of points that could have been allocated without determining show is $n^2-2n+4=2n+(n-2)(n-2)$ which is the number which would occur if there were 2 players with n points each and $n-2$ players tied at $n-2$ points each. However, the actual number of points two games beyond the height of the win tree is

$$(n^2-2n+1)+2+2=n^2-2n+5.$$

Since the maximum possible number of points is less than this actual number of points, show must have been determined two games beyond the height of the win tree, if not before. The examples, which show that the height of the extended tree is in fact $(n^2-n+4)/2$, are obvious extensions of the ones given above.

The win tree corresponding to the case $n=4$ is given in Fig. 1.

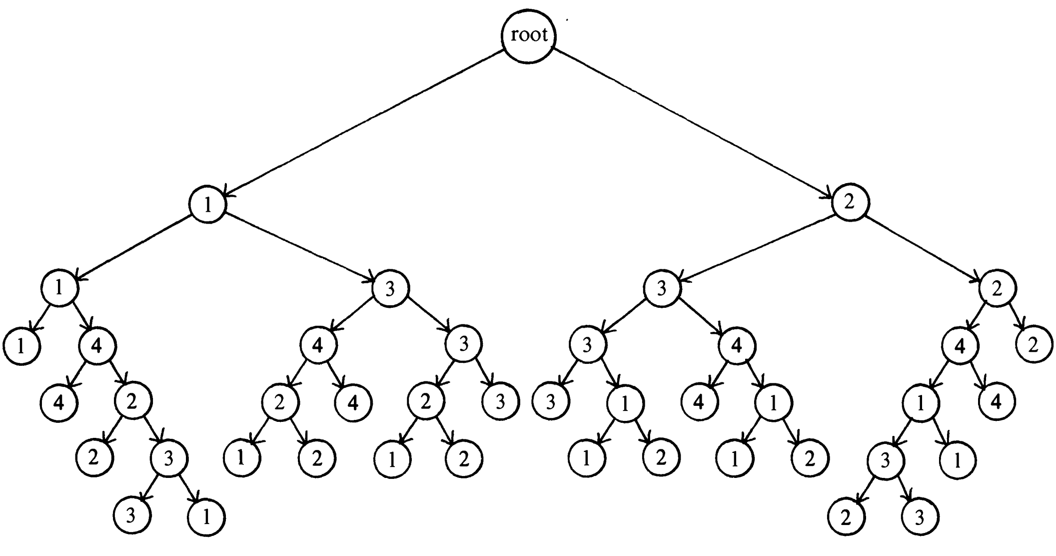


FIG. 1

By traversing the left subtree we obtain the following 11 possible game outcomes in which post position one wins the first match: 111, 1144, 11422, 114233, 114231, 1333, 13322, 13321, 1344, 13422, 13421. Similarly, by traversing the right subtree we obtain 11 possible game outcomes in which post position two wins the first match. This win tree is simpler than the extended tree, which includes branches from the win nodes for the playoffs for place and show, and is much simpler than the win tree for the standard game with eight players.

To generate the probabilities for win, place and show, as well as for the exacta and the quiniela, a computer program was written in the Pascal language; this language was chosen because of its recursive features. The program includes a recursive procedure which calls itself for each of the left and right subtrees, according to the recursive definition of preorder. For each of the various numbers of players or teams that might be tied for place or show, a separate procedure is included which is called for a playoff which involves that number of players or teams.

The recursive procedure is called in the main program twice, once each for the players in post positions one and two, depending on which of these two players wins the initial match. The procedures for the playoffs with five and seven players also each call the recursive procedure twice, for each of the left and right subtrees depending on which of the two players to play first in the playoff wins the match.

4. Abilities and Probabilities. In our initial version of the program we assume that the players are of equal abilities, that is, we assume that the probability of each player winning a match against any other player is 0.5. We assume further that the matches are independent events, that is, that the probability of a player in post position i winning against a player in post position j does not affect the probability of his subsequently winning against a player in post position k (we do not take into account such factors as fatigue, for example).

The assumption that each of the players is equally likely to win makes the computations considerably simpler. Among the reasons for this are the following: (1) The probability of reaching a node of height k is just $(0.5)^k$. (2) The players in post positions one and two have an equal chance of winning; hence, the results for the right subtree can be obtained from the results for the left subtree just by interchanging post positions one and two. (3) The probabilities in the playoff procedures can be easily calculated.

The computer time required to do the analysis for an eight-player game in which the players are of equal abilities for the first version of the program was 3.9 minutes.

In the second version of the program we deal with the more realistic situation in which the players are not necessarily of equal abilities. To each pair of post positions (i, j), $i \neq j$, we assign a probability p_{ij} that a player in post position i wins a match against a player in post position j. We set $p_{ii} = 0.5$ (though these values are not used in the computations). Thus we obtain the probability matrix (p_{ij}) where $p_{ji} = 1 - p_{ij}$ and $0 \leqslant p_{ij} \leqslant 1$.

Again we assume that the matches are independent events and label the edges of a directed binary tree with these probabilities. Then the probability of a certain game outcome, that is, the probability of reaching a certain node of the tree, is the product of the probabilities on the edges of the path from the root of the tree to that node, and the probability of a player in a certain post position winning a game is the sum of the probabilities at all the different leaves at which he is a winner. The place and show probabilities are determined similarly except that the extended tree must be considered.

For the exacta we consider the various combinations of post positions as ordered pairs (i, j), post position i to take first and post position j to take second, and let e_{ij} be the probability that (i, j) takes the exacta. For the quiniela the pairs are unordered so we denote these combinations by $\{i, j\}$, and let q_{ij} be the probability that $\{i, j\}$ takes the quiniela. An exacta probability e_{ij} cannot be obtained simply by multiplying the final probabilities for a player in post position i to win and a player in post position j to finish second. This exacta probability is a conditional probability that a player in post position j finishes second given that a player in post position i finishes first in a number of specific individual games. The exacta probability for an individual game is the product of the probabilities down to a leaf of the place tree and is the same as the probability of a player finishing second at that leaf. Each of the exacta probabilities e_{ij} must be computed by summing up the probabilities for the individual games in which the pair (i, j) takes the exacta. The quiniela probabilities, on the other hand, are found very simply by $q_{ij} = e_{ij} + e_{ji}$.

The analysis for an eight-player game in which the players are of equal abilities required 16.11 minutes of computer time for this second version of the program.

5. Results for Equal Abilities. We begin with the case in which the players are of equal ability. If the game were fair, the probabilities for each of the players to win, place and show would be 0.125, 0.25 and 0.375, whereas the exacta and quiniela probabilities for each of the pairs of players would be 0.0179 ($= 1/56$) and 0.0357 ($= 1/28$), respectively. As we see from Tables 1 and 2, these are not the values computed.

TABLE 1

Probabilities
Equal Abilities

Post Position	Win	Place	Show
1	0.1631	0.3404	0.4926
2	0.1631	0.3404	0.4926
3	0.1386	0.3035	0.4485
4	0.1240	0.2583	0.3937
5	0.1020	0.2112	0.3410
6	0.1026	0.1818	0.2939
7	0.0888	0.1718	0.2628
8	0.1177	0.1927	0.2750

First note that the players in post positions one and two have equal chances to win, place and show, as they must if they have equal abilities. At 0.1631 the win probability for a player in post position one or two is 0.0381 higher than the average win probability (0.125). Furthermore, the win probability for a player in post position one or two is nearly twice that of a player in post position seven, whose win probability is only 0.0888. The probability (0.1177) of a player's winning in post position eight is higher than the probabilities of his winning in post positions five, six and especially seven. Indeed, the win probability for a player in post position eight is almost as high as the win probability for a player in post position four (0.1240).

Note also that the place and show probabilities are not just two and three times the win probabilities. In the case of equal abilities the highest probabilities for place and show occur for the players in post positions one and two. The place and show probabilities for a player in post position three are also comparatively high. The lowest probabilities for place and show occur for the players in post positions six, seven and eight.

TABLE 2

Exacta Probabilities
Equal Abilities

First \ Second	1	2	3	4	5	6	7	8
1		0.0327	0.0280	0.0226	0.0228	0.0194	0.0199	0.0177
2	0.0327		0.0280	0.0226	0.0228	0.0194	0.0199	0.0177
3	0.0337	0.0337		0.0154	0.0126	0.0110	0.0155	0.0168
4	0.0331	0.0331	0.0226		0.0088	0.0058	0.0102	0.0103
5	0.0244	0.0244	0.0234	0.0132		0.0049	0.0062	0.0056
6	0.0212	0.0212	0.0241	0.0188	0.0085		0.0055	0.0033
7	0.0149	0.0149	0.0181	0.0187	0.0131	0.0054		0.0035
8	0.0173	0.0173	0.0207	0.0229	0.0205	0.0133	0.0057	

Note furthermore that the exacta probabilities are not symmetric about the diagonal. In the

case of equal abilities the exacta probabilities are highest for the combinations (3, 1), (3, 2), (4, 1), (4, 2), (1, 2) and (2, 1). The quiniela probabilities q_{ij}, which may be found by adding the exacta probabilities e_{ij} and e_{ji}, are highest for the combinations $\{1,2\}$, $\{1,3\}$ and $\{2,3\}$.

6. Results with Handicapping. When one looks at the historical record and compares the statistics from games actually played over a period of time at various frontons with the figures in Table 1, one finds that the percentages of the time the players in the various post positions actually win are usually more uniform than the computed probabilities. A reason for this is that the house usually handicaps the better players by starting them in the less favorable post positions in order to obtain a more uniform distribution of probabilities for each of the post positions to win.

When such handicapping is done to equalize the probabilities for win, one cannot necessarily expect that the probabilities for place and show will be evened out simultaneously. To investigate the effect of handicapping on the place and show probabilities, we experimentally determined a matrix of abilities which gives an approximately uniform distribution of win probabilities. Relatively slight adjustments in the abilities had to be made to get a uniform outcome for win. The matrix, which is given below, is not unique, of course.

Ability Matrix

Player	1	2	3	4	5	6	7	8
1	0.50	0.50	0.48	0.45	0.42	0.43	0.42	0.47
2	0.50	0.50	0.48	0.45	0.42	0.43	0.42	0.47
3	0.52	0.52	0.50	0.48	0.46	0.46	0.46	0.50
4	0.55	0.55	0.52	0.50	0.47	0.48	0.47	0.51
5	0.58	0.58	0.54	0.53	0.50	0.50	0.50	0.53
6	0.57	0.57	0.54	0.52	0.50	0.50	0.49	0.52
7	0.58	0.58	0.54	0.53	0.50	0.51	0.50	0.52
8	0.53	0.53	0.50	0.49	0.47	0.48	0.48	0.50

The probabilities for win, place and show corresponding to this ability matrix are given in Table 3.

TABLE 3

Probabilities
Nearly Equal Win Probabilities

Post Position	Win	Place	Show
1	0.1256	0.2918	0.4426
2	0.1256	0.2918	0.4426
3	0.1250	0.2905	0.4386
4	0.1228	0.2617	0.4018
5	0.1258	0.2439	0.3801
6	0.1260	0.2056	0.3140
7	0.1237	0.2149	0.3025
8	0.1255	0.1995	0.2773

Notice that the place and show probabilities are not uniform. Even though the probability of winning is almost uniform, for place and show there is still a strong advantage for the players in post positions one, two and three, while the players in post positions six, seven and eight remain at a disadvantage. Although we do not present the table here, we have found further that the exacta probabilities are highest for the combinations (4, 1), (4, 2), (6, 3) and (5, 3) and that the quiniela probabilities are highest for the combinations $\{1,5\}$, $\{2,5\}$, $\{1,4\}$, $\{2,4\}$, $\{1,3\}$ and $\{2,3\}$.

7. Results for a World Class Player. Of course, there are situations in which one player, a world class player, is better than all the other players. Considering this possibility, we assume that the probability of such a world class player's winning a match against each of the other players is 0.67 (it is rare for the world class player's probability to be greater than 0.67) and that all the other players are of equal abilities. We assume further that the handicappers place this world class player in post positions five, six, seven or eight. The handicappers would not usually place the world class player in post positions one, two, three or four since, as we have seen (Table 1), these post positions have an intrinsic advantage. The results of the analysis in each of these cases are given in Table 4.

TABLE 4

Probabilities

World Class Player in Post Position

	Five			Six			Seven			Eight		
	Win	Place	Show	Win	Place	Show	Win	Place	Show	Win	Place	Show
1	0.1329	0.3122	0.4724	0.1368	0.3166	0.4683	0.1407	0.3083	0.4554	0.1323	0.2966	0.4414
2	0.1329	0.3122	0.4724	0.1368	0.3166	0.4683	0.1407	0.3083	0.4554	0.1323	0.2966	0.4414
3	0.1045	0.2699	0.4261	0.1090	0.2843	0.4294	0.1186	0.2830	0.4256	0.1166	0.2801	0.4200
4	0.0919	0.1989	0.3394	0.0888	0.2225	0.3674	0.0986	0.2411	0.3802	0.1007	0.2458	0.3749
5	0.2869	0.4612	0.6078	0.0731	0.1656	0.2863	0.0744	0.1776	0.3186	0.0760	0.1944	0.3247
6	0.0773	0.1398	0.2280	0.2927	0.4151	0.5498	0.0735	0.1331	0.2398	0.0697	0.1477	0.2753
7	0.0721	0.1405	0.2152	0.0660	0.1305	0.2027	0.2643	0.4004	0.5125	0.0603	0.1195	0.2068
8	0.1015	0.1654	0.2392	0.0969	0.1588	0.2273	0.0891	0.1482	0.2123	0.3119	0.4194	0.5152

As one would expect, the probabilities for the world class player to win, place and show in any of these positions are much higher than those of the other players. From Table 4 we see that the world class player's win probability (0.3119) is highest when he is in post position eight, and that his highest place (0.4612) and show (0.6078) probabilities occur when he is in post position five. In post position seven his probabilities are not quite as high as when he is in the other post positions. Although the probability of the world class player's winning is more than twice as great as that of the players in post positions one and two, these players still enjoy a slight margin to win and quite substantial margins to place and show over the other players. The players in post positions one or two have the best chances to win when the world class player is in post position seven, to place when he is in post position six and to show when he is in post position five.

Further, we have found that the highest exacta and quiniela probabilities occur when the world class player is in post position five rather than six, seven or eight. We have also found that, if the world class player is in post position five or six, the highest exacta probability occurs for the world class player taking first and the player in post position three taking second, whereas if the world class player is in post position seven or eight, the highest exacta probability occurs for the world class player taking first and the player in post position four taking second. Likewise, for the quiniela we have found that if the world class player is in post position five or six, the highest probability occurs for the world class player and the player in post position one or two, whereas if the world class player is in post position seven or eight, the highest quiniela probability occurs for the world class player and the player in post position three.

8. Number of Different Game Outcomes. For the standard eight player game of Jai Alai the total number of different ways of choosing a winner (in other words, the number of leaves of the win tree) was determined to be 422,384. A computer count of the numbers of games which involve

playoffs for place and/or show was also done. These figures are presented in Table 5.

TABLE 5

Games in Which a Playoff is Involved

Number of Players or Teams Tied	Place	Show
1	102,302	35,948
2	132,028	29,078
3	109,468	21,538
4	57,276	8,826
5	18,022	6,468
6	2,962	444
7	326	

Note that one player or team tied for place or show means that there is no playoff, that place or show is automatically awarded to the player with the second or third highest number of points. In the second column the 422,384 games are broken down into games which involve one to seven players tied for place; the 102,302 games in which one person is automatically awarded place are further broken down in the third column into games in which one to six players are tied for show. Approximately one-fourth of the games involve no playoff for place, and one-twelfth of the games involve no playoff for place or show. A seven-way tie for place occurs in less than one out of one thousand games.

The results presented in this paper are of great importance to Jai Alai gamblers. Because a bet is not just a bet on a certain post position but a bet against the other gamblers, a serious gambler must take into account the pari-mutuel odds and the available information on each player or team in conjunction with the figures we have calculated. What a gambler does with these figures, given particular circumstances and odds, is his business.

References

1. State of Florida Administrative Code, Chapter 7E-3, Jai Alai Rules and Regulations, August 1974.
2. State of Nevada Gaming Commission Regulation 27, Jai Alai, March 1975.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

ON THE MEAN VALUE THEOREM FOR INTEGRALS

BERNARD JACOBSON
Department of Mathematics, Franklin and Marshall College, Lancaster, PA 17604

The Mean Value Theorem for Integrals states:

THEOREM 1. *If f is a continuous function for $a \leqslant t \leqslant x$, then there exists a number c such that $a < c < x$ and $\int_a^x f(t)\,dt = f(c)(x-a)$.*

Falling into either the category of "interesting facts we once knew but have forgotten" or the

category of "interesting facts we should have known but never did" is the fact that as x approaches a the value of c approaches the midpoint between a and x.

An unscientific survey of mathematicians at several institutions leads me to believe that this fact is generally unknown. The result and a short proof are given below.

THEOREM 2. *If f is differentiable at a, $f'(a) \neq 0$, and c is taken as in Theorem 1, then*

$$\operatorname{limit}_{c\to a} \frac{c-a}{x-a} = \frac{1}{2}.$$

Proof. Consider

$$\text{(A)} \qquad \operatorname{limit}_{c\to a} \frac{\int_a^x f(t)\,dt - xf(a) + af(a)}{(x-a)^2}.$$

Upon our applying the Mean Value Theorem

$$\operatorname{limit}_{c\to a} \frac{\int_a^x f(t)\,dt - xf(a) + af(a)}{(x-a)^2} = \operatorname{limit}_{c\to a} \frac{f(c)(x-a) - f(a)(x-a)}{(x-a)^2} = \operatorname{limit}_{c\to a} \frac{f(c)-f(a)}{x-a}$$

$$= \operatorname{limit}_{c\to a} \frac{f(c)-f(a)}{c-a} \cdot \frac{c-a}{x-a} = f'(a) \cdot \operatorname{limit}_{c\to a} \frac{c-a}{x-a}.$$

Applying L'Hospital's rule to (A) we obtain:

$$\operatorname{limit}_{c\to a} \frac{\int_a^x f(t)\,dt - xf(a) + af(a)}{(x-a)^2} = \operatorname{limit}_{c\to a} \frac{f(x)-f(a)}{2(x-a)} = \frac{f'(a)}{2}.$$

Thus

$$\operatorname{limit}_{c\to a} \frac{c-a}{x-a} = \frac{1}{2},$$

which is the result of Theorem 2.

GAMMA FUNCTION DERIVATION OF n-SPHERE VOLUMES

GREG HUBER

Department of Mathematics, Massachusetts Institute of Technology, Room 24-612, Cambridge, MA 02139

Introduction. Beginning students of analysis are often presented with a simple inductive derivation of n-sphere volume formulas. This induction leads to an expression of the form $f(n)/\Gamma(g(n))$, where f and g are some functions of n. In this short note we look at the gamma function as a starting point in an alternative derivation. As a preliminary, some of the basic properties of the gamma function would be useful.

Preliminaries. By definition, $\Gamma(p) = \int_0^\infty t^{p-1}e^{-t}\,dt$. We shall be interested in the cases where $t = a^2$, so that

$$\Gamma(p) = \int_0^\infty (a^2)^{p-1} e^{-a^2}\,d(a^2) = 2\int_0^\infty a^{2p-1}e^{-a^2}\,da.$$

Letting $p = \frac{1}{2}$ yields

$$\Gamma(\tfrac{1}{2}) = 2\int_0^\infty e^{-a^2}\,da = \int_{-\infty}^\infty e^{-a^2}\,da = \sqrt{\pi}.$$

Volume elements. To S^{n-1}, we associate a volume $V_n(S^{n-1})$ over $\{x: \|x\| \leqslant 1\}$, and a

"surface" volume $\omega_{n-1}(S^{n-1})$ over S^{n-1}, where $x \in R^n$ and $\|x\| = \sqrt{\Sigma_{i=0}^{n-1} x_i^2}$, the usual norm on R^n. (The symbols V_n and ω_{n-1} will be substituted for $V_n(S^{n-1})$ and $\omega_{n-1}(S^{n-1})$ for brevity.) The two are related by the integral

$$V_n = \int_0^1 \omega_{n-1} r^{n-1}\, dr = \omega_{n-1}/n.$$

However, we may compare this with the volume element $\Pi_{i=0}^{n-1} dx_i$ and write

$$V_n = \int_{S^{n-1}} \cdots \int \Pi_{i=0}^{n-1} dx_i = \int_0^1 \omega_{n-1} r^{n-1}\, dr \tag{1}$$

where $r = \|x\| = \sqrt{\Sigma_i x_i^2}$.

Integrating e^{-r^2} over all of R^n using one volume element (given on the right side of (1)) yields

$$\int_0^\infty \omega_{n-1} r^{n-1} e^{-r^2}\, dr = \omega_{n-1} \int_0^\infty r^{n-1} e^{-r^2}\, dr = \tfrac{1}{2}\omega_{n-1}\Gamma\left(\tfrac{1}{2}n\right). \tag{2}$$

Now, using the left side of (1), we get

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-r^2} \Pi_i\, dx_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Pi_i\left(e^{-x_i^2}\, dx_i\right) = \Pi_i \int_{-\infty}^{\infty} e^{-x_i^2}\, dx_i$$
$$= \left(\int_{-\infty}^{\infty} e^{-x^2}\, dx\right)^n = \pi^{1/2n}.$$

Equating this with (2) gives

$$\omega_{n-1}(S^{n-1}) = 2\pi^{1/2n}/\Gamma\left(\tfrac{1}{2}n\right).$$

We saw before that $V_n = \omega_{n-1}/n$; therefore

$$V_n(S^{n-1}) = 2\pi^{1/2n}/\left(n\Gamma\left(\tfrac{1}{2}n\right)\right) = \pi^{1/2n}/\left(\tfrac{1}{2}n\Gamma\left(\tfrac{1}{2}n\right)\right)$$
$$= \pi^{1/2n}/\Gamma\left(1 + \tfrac{1}{2}n\right),$$

and we have the desired expressions.

A CONVERSE TO ROUCHÉ'S THEOREM

DAVID CHALLENER AND LEE RUBEL
Department of Mathematics, University of Illinois, Urbana, IL 61801

The purpose of this note is to point out a converse to Rouché's Theorem as strengthened by T. Estermann in **[2]** and independently by I. Glicksberg in **[3]**. Assume f, g are analytic in $|z| \leqslant 1 + \varepsilon$, having no zeros on $|z| = 1$. Let Z_f, Z_g denote the number of zeros of f, g in $|z| < 1$. Estermann proved that $|f + g| < |f| + |g|$ on $|z| = 1$ is a sufficient condition that $Z_f = Z_g$. A simple argument then shows that $Z_f = Z_g$ if there exist finite Blaschke products α, β *of the same order* such that $|\alpha f + \beta g| < |f| + |g|$ on $|z| = 1$. We show here that this condition is also necessary for $Z_f = Z_g$. Recall that a finite Blaschke product is a function of the form:

$$\alpha(z) = e^{i\lambda} z^m \prod_{i=1}^{n} \frac{-|a_i|(z - a_i)}{a_i(1 - \bar{a}_i z)} \qquad m \geqslant 0 \quad \text{and} \quad 0 < |a_i| < 1,$$

where λ is an arbitrary real constant. Note that each factor is a Möbius function that maps the unit disc 1 : 1 and onto the unit disc. Hence $|z| = 1$ implies $|\alpha(z)| = 1$. For the purposes of this paper we allow $\alpha(z) = e^{i\lambda} z^m$ to be considered a finite Blaschke product. The order of a finite Blaschke product is $m + n$, which is the number of zeros of $\alpha(z)$ inside the unit disc.

The following example shows that Estermann's result as it stands does *not* admit a converse: Let G be the region in Fig. 1. Then by the Riemann Mapping Theorem, there exists an analytic homeomorphism $h:\mathbb{D}=\{|z|<1\}\to G$. By Carathéodory's result in [5], h extends to be continuous on $|z|\leq 1$ and $h|_{\partial\mathbb{D}}$ is a homeomorphism from $\partial\mathbb{D}$ to ∂G. Let $\delta>0$. Then there exists $\varepsilon>0$ such that

$$\max_{0\leq\theta<2\pi}\left|h\left((1-\varepsilon)e^{i\theta}\right)-h\left(e^{i\theta}\right)\right|<\delta.$$

Let $G_\delta=h\{|z|<1-\varepsilon\}$. Then $H(z)=h((1-\varepsilon)z)$ maps $\mathbb{D}$ to G_δ homeomorphically and has the following properties:

(i) $H(z)$ is analytic in $|z|<1+\varepsilon$.
(ii) $H(z)$ has no zeros in $|z|\leq 1$.
(iii) For δ sufficiently small, $\arg[H(z)]$ assumes all values a with $0\leq a<2\pi$ on $|z|=1$.

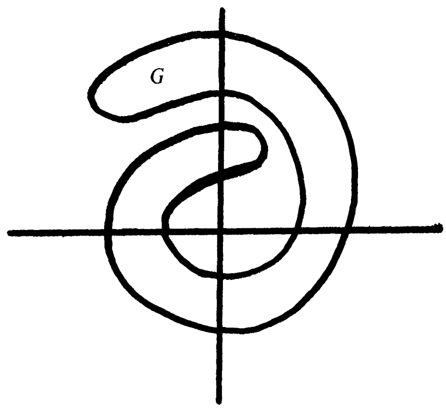


FIG. 1

Yet if $f(z)\equiv 1$ and $e^{i\theta}$ is any complex number of modulus one, there is a z with $|z|=1$ such that $|e^{i\theta}f(z)+H(z)|=|f(z)|+|H(z)|$. This fact shows that Estermann's result does not admit a converse even if we are allowed to rotate $f(z)$ arbitrarily.

Our proof rests on the following fact proved by Helson and Sarason in [4].

FACT: *Any continuous unimodular function* γ *on* $|z|=1$ *is the uniform limit of quotients of finite Blaschke products.*

The proof we give of this is due to Davie, Gamelin, and Garnett in [1].

Sketch of Proof. Let $\gamma(z)=z^m v^2$ for some continuous unimodular function v defined on $|z|=1$. Then v can be uniformly approximated by a rational function h, using the Cesàro means of the Fourier series for v. To complete the proof, observe that, on $|z|=1$,

$$\begin{aligned} z^m h(z)/\overline{h(1/\bar{z})} &= z^m h(z)/\overline{h(z)} \\ &= z^m h^2(z)/|\overline{h(z)}|^2, \end{aligned}$$

which, as v is unimodular, uniformly approximates $z^m h^2(z)$. Further we note that $z^m h(z)/\overline{h(1/\bar{z})}$ is a finite Blaschke product.

LEMMA 1. *Given* f, g *analytic in* $|z|\leq 1+\varepsilon$, $\varepsilon>0$, f *and* g *having no zeros on* $|z|=1$, *then there exist finite Blaschke products* α, β *such that* $|\alpha f+\beta g|<\max\{|f|,|g|\}$ *on* $|z|=1$.

Proof. Consider $h = g/f$ on $|z| = 1$. Define

$$\gamma(e^{i\theta}) = \frac{h(e^{i\theta})}{|h(e^{i\theta})|}.$$

Then γ is a continuous unimodular function on $|z| = 1$, as g and f are never zero and are continuous on $|z| = 1$. Then, by **[1]**, there exist finite Blaschke products α, β, such that

$$|-\gamma - \alpha/\beta| < \frac{\min\limits_{|z|=1}\{|g|,|f|\}}{\max\limits_{|z|=1}|f|} \text{ on } |z| = 1.$$

Thus, on $|z| = 1$,

$$|\beta g + \alpha f| = \left|\frac{g}{f} + \frac{\alpha}{\beta}\right||f| = \left|h + \frac{\alpha}{\beta}\right||f| = \left||h|\gamma + \frac{\alpha}{\beta}\right||f|$$

$$\leqslant |f|\left[\left|\gamma + \frac{\alpha}{\beta}\right| + ||h| - 1|\right] < |f|\left|\frac{\min\limits_{|z|=1}\{|g|,|f|\}}{\max\limits_{|z|=1}|f|} + ||h| - 1|\right|$$

$$\leqslant \min_{|z|=1}\{|g|,|f|\} + ||g| - |f|| \leqslant \max\{|g|,|f|\}.$$

Therefore $|\beta g + \alpha f| < \max\{|g|,|f|\}$ on $|z| = 1$.

THEOREM 1. *Suppose f and g are analytic in $|z| \leqslant 1 + \varepsilon$, $\varepsilon > 0$, with no zeros on $|z| = 1$. If Z_f, Z_g are the number of zeros of f and g in $|z| < 1$, counted according to their multiplicities, then: $Z_f = Z_g$ if and only if there exist finite Blaschke products α, β of the same order such that $|\alpha f + \beta g| < |f| + |g|$ on $|z| = 1$.*

Proof. $\Rightarrow$ Assume $Z_f = Z_g$. By Lemma 1, there exist α, β, Blaschke products of finite order, such that $|\alpha f + \beta g| < |f| + |g| = |\alpha f| + |\beta g|$ on $|z| = 1$. Therefore, by Estermann's strong version of Rouché's Theorem **[2]**, αf and βg have the same number of zeros; i.e., $Z_{\alpha f} = Z_{\beta g}$. Thus $Z_\alpha + Z_f = Z_\beta + Z_g$. By our assumption, then $Z_\alpha = Z_\beta$.

$\Leftarrow$ Assume there exist finite Blaschke products α, β of the same order such that $|\alpha f + \beta g| < |f| + |g|$. Then $|\alpha f + \beta g| < |\alpha f| + |\beta g|$ on $|z| = 1$, so that by Rouché's Theorem **[2]**, $Z_\alpha + Z_f = Z_\beta + Z_g$. But, by hypothesis, $Z_\alpha = Z_\beta$. Thus $Z_f = Z_g$.

COROLLARY 1. *Suppose that f, g are analytic in $|z| \leqslant 1 + \varepsilon$, that $\varepsilon > 0$, and that f and g have no zeros on $|z| = 1$. Then $Z_f - Z_g = m$ if and only if there are finite Blaschke products α, β such that $Z_\beta - Z_\alpha = m$ and $|\alpha f + \beta g| < |f| + |g|$ on $|z| = 1$.*

We omit the proof, which is almost the same as the proof of our theorem.

COROLLARY 2. *If there are finite Blaschke products α, β such that $Z_\alpha = Z_\beta$ and $|\alpha f + \beta g| < |f| + |g|$ on $|z| = 1$, then there exist finite Blaschke products α^*, β^* such that $Z_{\alpha^*} = Z_{\beta^*}$ and $|\alpha^* f + \beta^* g| < \max\{|f|,|g|\}$.*

Proof. By Theorem 1, $Z_\alpha = Z_\beta$ implies $Z_f = Z_g$. Also, by Lemma 1, there exist α^*, β^*, finite Blaschke products, such that $|\alpha^* f + \beta^* g| < \max\{|f|,|g|\}$. But then, by Theorem 1, $Z_f = Z_g$ implies $Z_{\alpha^*} = Z_{\beta^*}$.

Further we note that Lemma 1, and hence our converse, is true for f, g in the disc algebra, i.e., functions continuous for $|z| \leqslant 1$, and analytic for $|z| < 1$, provided that f, g are nonzero on the boundary $|z| = 1$. Our results also extend in the obvious way for f, g meromorphic.

The work of the second author was partially supported by a grant from the National Science Foundation.

References

1. A. M. Davie, T. W. Gamelin and J. Garnett, Distance estimates and pointwise bounded density, Trans. Amer. Math. Soc., 175 (1973) 62.

2. T. Estermann, Complex Numbers and Functions, The Athlone Press, University of London, London, 1962, Theorem 18.2, p. 157.

3. Irving Glicksberg, A remark on Rouché's Theorem, this MONTHLY, 83 (1976) 186–187.

4. H. Helson and D. Sarason, Past and future, Math. Scand., 21 (1967) Lemma on page 9.

5. A. I. Markushevich, Theory of Functions of a Complex Variable, vol. 3, Prentice-Hall, Englewood Cliffs, N.J., 1976, Theorem 2.24, p. 70.

A REFLECTION PRINCIPLE FOR CALCULATING THE DERIVATIVES OF A POLYNOMIAL

PABLO M. SALZBERG

Universidad Simón Bolívar, Departamento de Matemáticas, Caracas, Venezuela

In what follows, given a homogeneous polynomial of degree m, we exhibit a fairly simple algorithm for calculating the derivatives of order $m-k$ from those of order k, without performing any further differentiation.

To this effect, given a polynomial $H(Y)$ in $Y=(Y_1,\dots,Y_n)$, and an auxiliary vector $X=(X_1,\dots,X_n)$, we define recursively:

$$\begin{aligned}\nabla^1 H(Y)\cdot X^1 &:= \nabla H(Y)\circ X,\\ \nabla^k H(Y)\cdot X^k &:= \nabla\left(\nabla^{k-1}H(Y)\cdot X^{k-1}\right)\circ X,\end{aligned} \tag{1}$$

where "∇" denotes the gradient operator and "$\circ$" denotes the scalar product.

Therefore, if we consider for example $H(Y)=H(Y_1,Y_2)=Y_1^3-Y_1^2Y_2$ and take $X=(X_1,X_2)$ as an auxiliary vector, then we have:

$$\begin{aligned}\nabla^1 H(Y)\cdot X^1 &= \nabla H(Y_1,Y_2)\circ(X_1,X_2)=\left(3Y_1^2-2Y_1Y_2,-Y_1^2\right)\circ(X_1,X_2)\\ &=\left(3Y_1^2-2Y_1Y_2\right)X_1-Y_1^2X_2;\\ \nabla^2 H(Y)\cdot X^2 &=\left[\nabla\left(3Y_1^2X_1-2Y_1Y_2X_1-Y_1^2X_2\right)\right]\circ(X_1,X_2)\\ &=(6Y_1X_1-2Y_2X_1-2Y_1X_2,-2Y_1X_1)\circ(X_1,X_2)\\ &=6Y_1X_1^2-2Y_2X_1^2-4Y_1X_1X_2;\text{ and so on.}\end{aligned}$$

We may now state the following *reflection principle* for the derivates of any homogeneous polynomial H of degree m: *The coefficients of $\nabla H(Y)\circ X$ regarded as a polynomial in Y are, up to constants, the derivatives of order $m-1$ of H. Analogously, the coefficients of $\nabla(\nabla H(Y)\circ X)\circ X$ are, up to constants, the derivatives of order $m-2$ of H, and so forth. More explicitly,*

$$\frac{\partial^{|s|}H(X_1,\dots,X_n)}{\partial X_1^{s_1}\cdots\partial X_n^{s_n}}$$

(where $|s|=\sum_{i=1}^n s_i$) is given by the coefficient of the monomial in $Y_1^{s_1}\cdots Y_n^{s_n}$ of $\nabla^k H(Y)\cdot X^k$ times $s_1!s_2!\cdots s_n!/k!$, where $k=m-|s|$.

Therefore, to calculate

$$\frac{\partial^{|s|}H(X_1,\dots,X_n)}{\partial X_1^{s_1}\cdots\partial X_n^{s_n}}$$

we must perform the following steps:

1. calculate $\nabla^k H(Y) \cdot X^k$ as indicated by (1), where $k = m - |s|$;
2. display $\nabla^k H(Y) \cdot X^k$ as a polynomial in Y (by reordering the terms in powers of Y);
3. look at the coefficients of $Y_1^{s_1} \cdots Y_n^{s_n}$;
4. take $\dfrac{s_1! \cdots s_n!}{k!}$ times this coefficient.

EXAMPLE: Let $H(X, Y, Z) = X^8 + 3XY^7 - 2X^2Y^6 + X^3Z^5$.

To calculate, for instance, the derivatives of order 7, we must perform the following manipulations:

1. We form the scalar product between the gradient of H and an auxiliary vector, say (x, y, z):

$$\nabla H(X, Y, Z) \cdot (x, y, z) = (8X^7 + 3Y^7 - 4XY^6 + 3X^2Z^5)x + (21XY^6 - 12X^2Y^5)y + (5X^3Z^4)z.$$

2. By reordering in powers of X, Y, Z we obtain:

$$\nabla H(X, Y, Z) \cdot (x, y, z) = 8xX^7 + 3xY^7 + (-4x + 21y)XY^6 - 12yX^2Y^5 + 3xX^2Z^5 + 5zX^3Z^4.$$

3. The coefficients of the above polynomial are, up to constants, the partial derivatives of order 7 of H.
4. Hence, we obtain

$$\frac{\partial^7 H}{\partial x^7} = 7!8x, \quad \frac{\partial^7 H}{\partial y^7} = 7!3x, \quad \frac{\partial^7 H}{\partial x \partial y^6} = 6!\,(-4x + 21y),$$

$$\frac{\partial^7 H}{\partial x^2 \partial y^5} = 2!5!12y, \quad \frac{\partial^7 H}{\partial x^2 \partial z^5} = 2!5!3x \quad \text{and} \quad \frac{\partial^7 H}{\partial x^3 \partial z^4} = 3!4!5z;$$

all the remaining derivatives are zero.

To prove this algorithm we shall need some additional notation. Given $s = (s_1, s_2, \ldots, s_n) \in N^n$, we set:

$$|s| := \sum_{i=1}^{n} s_i, \quad s! := s_1! \cdots s_n!, \quad Y^s := Y_1^{s_1} \cdots Y_n^{s_n}, \quad \text{and} \quad D_s H(Y) := \frac{\partial^{|s|} H(Y)}{\partial Y_1^{s_1} \cdots \partial Y_n^{s_n}}.$$

At this point the reader should observe that other notations for (1) are in use, namely $f^{(k)}(Y; X)$ [**1**, § 12.4] and $\left(\sum_{i=1}^n X_i \frac{\partial}{\partial Y_i}\right)^k H(Y)$ [**2**, p. 122]. Moreover, expanding the latter expression by means of the multinomial formula we obtain [**2**, p. 122]:

$$\left(\sum_{i=1}^{n} X_i \frac{\partial}{\partial Y_i}\right)^k H(Y) = \sum_{|r|=k} \frac{k!}{r_1! \cdots r_n!} X_1^{r_1} \cdots X_n^{r_n} \frac{\partial^k H(Y)}{\partial Y_1^{r_1} \cdots \partial Y_n^{r_n}};$$

or equivalently, in our notation:

$$\nabla^k H(Y) \cdot X^k = \Sigma_{|r|=k} \frac{k!}{r!} X^r D_r H(Y).$$

Now, the reflection principle follows in a straightforward manner from the following result:

THEOREM. *Let H be a homogeneous polynomial of order m. Then the following identities hold*:

$$\frac{1}{k!} \nabla^k H(Y) \cdot X^k = \sum_{|s|=m-k} \frac{1}{s!} D_s H(X) Y^s, \quad 0 \leq k \leq m. \tag{2}$$

Proof. The left-hand side in (2) is the kth homogeneous component of the Taylor expansion of $H(X+Y)$ centered at Y [**2**, p. 122]. Moreover, the right-hand side in (2) is the $(m-k)$th homogeneous component of the Taylor expansion of $H(X+Y)$ centered at X. As polynomials in

X, both sides are the kth homogeneous component of the same polynomial $H^*(X) = H(X+Y)$; hence, they must be equal.

Alternately, this result can be obtained directly using induction on $m - |s|$ and Euler's theorem on homogeneous functions [**1**, pp. 364–365].

REMARKS.

1. Let $P(X)$ be a polynomial and write $P(X) = \Sigma_{i=0}^{n} P_i(X)$ where P_i denotes the ith homogeneous component of P. Since $\nabla^k[\cdot] \cdot X^k$ is a linear operator, then for any monomial $c_s(X)Y^s$ of $\nabla^k[P(Y)] \cdot X^k$ (regarded as a polynomial in Y) the equality:
$$D_s P_{|s|+k}(X) = \frac{s!}{k!} c_s(X),$$
holds.
2. Let us write $\nabla^p H(Y) X^p = \Sigma_{|s|=m-p} c_s(X) Y^s$. To obtain any term $c_s(X)Y^s$ of $\nabla^k H(Y) \cdot X^k$ from $\nabla^{k-1} H(Y) \cdot X^{k-1}$, we need look only for the terms of the form $c_{s+u_i}(X) Y^{s+u_i}$, where $u_i = (0, \ldots, 1, \ldots, 0)$ denotes the ith coordinate vector.
Hence, the following equality holds:
$$c_s(X) = \sum_{i=1}^{n} (s_i + 1) c_{s+u_i}(X) \cdot X_i. \tag{3}$$

Thus, to calculate the derivative $\partial^7 H/\partial x \partial y^6$ in the example above, since $s = (1, 6, 0)$, we need look only for the coefficients of X^2Y^6, XY^7 and XY^6Z of $H(X, Y, Z)$.
Hence, by (3), $\partial^7 H/\partial x \partial y^6 = 6!(2(-2)x + 7 \cdot 3y)$.

The author wishes to thank York University, Ontario, Canada, for the hospitality extended to him during his sabbatical leave.

References

1. T. Apostol, Mathematical Analysis, 2nd ed., Addison-Wesley, Reading, Mass., 1975.
2. W. Hodge and D. Pedoe, Methods of Algebraic Geometry, vol. 1, Cambridge University Press, New York, 1952.

THE BOUNDARY TOPOLOGY OF A SPACE

BRIAN M. SCOTT
Department of Mathematics, Cleveland State University, Cleveland, OH 44115

ZACHARY ROBINSON
Room J-117, MacGregor Dormitory, 450 Memorial Drive, Cambridge, MA 02139

There are many examples in general topology of the phenomenon of one topology on a set being defined in terms of another. For example, given a space $\langle X, \mathfrak{T} \rangle$ one may form, *inter alia*, the G_δ-topology, $\mathfrak{T}_\delta$, which has as a base all G_δ-sets of X with respect to $\mathfrak{T}$; the sequential topology, $\mathfrak{T}_s$, given as $\{V \subseteq X\colon X \setminus V$ contains the limits of all its $\mathfrak{T}$-convergent sequences$\}$; and the semi-regularization of $\mathfrak{T}$, $\mathfrak{T}_{\mathrm{sr}}$, generated by the base $\{\mathrm{int}_{\mathfrak{T}} \mathrm{cl}_{\mathfrak{T}} V \colon V \in \mathfrak{T}\}$ of regularly $\mathfrak{T}$-open sets. However, this mode of getting new spaces from old is generally ignored in undergraduate texts, presumably as being too sophisticated. In this note we present a simple example of this idea dealing with a basic concept of topology, *viz.*, the boundary operator. We first characterize those subsets of a space $\langle X, \mathfrak{T} \rangle$ that are boundaries and then prove that they (together with the space itself) are precisely the closed sets of a coarser topology, $\mathfrak{T}^{\#}$, on X. We then show that iterating this process gives nothing new, determine when $\mathfrak{T}^{\#} = \mathfrak{T}$, and mention some related results.

Let $\langle X, \mathfrak{T} \rangle$ be a topological space, and let $\mathfrak{B} = \{\mathrm{bdry}\, A : A \subseteq X\}$. Call a set $A \subseteq X$ resolvable

if and only if there are disjoint sets D, $E \subseteq A$, both of which are dense in A. ($\varnothing$ is vacuously resolvable.)

LEMMA 1. *Let $A \subseteq X$: then $A \in \mathfrak{B}$ if and only if* (i) *A is closed and* (ii) int *A is resolvable.*

Proof. Suppose that A is closed and that $D \subseteq \operatorname{int} A$ is such that D and $(\operatorname{int} A)\setminus D$ are dense in $\operatorname{int} A$; direct computation shows that $A = \operatorname{bdry}(A\setminus D)$.

If, now, $A = \operatorname{bdry} S$, then certainly A is closed, and it is easy to see that $S \cap \operatorname{int} A$ and $(\operatorname{int} A)\setminus S$ must both be dense in $\operatorname{int} A$. ■

$$\text{Let } \mathfrak{T}^* = \{X\setminus B : B \in \mathfrak{B}\}, \text{ and let } \mathfrak{T}^\# = \mathfrak{T}^* \cup \{\varnothing\}.$$

THEOREM 1. (i) *$\mathfrak{T}^\#$ is a topology on X.* (ii) $(\mathfrak{T}^\#)^\# = \mathfrak{T}^\#$.

Proof. Throughout the proof any topology-dependent assertion (such as '$A \subseteq X$ is resolvable') refers to the topology $\mathfrak{T}$ unless otherwise qualified.

(i) It suffices to verify that $\mathfrak{B} \cup \{X\}$ contains $\varnothing$ and X and that $\mathfrak{B}$ is closed under finite unions and arbitrary intersections. Obviously $X \in \mathfrak{B} \cup \{X\}$; and $\varnothing = \operatorname{bdry} \varnothing \in \mathfrak{B}$. Suppose that $\varnothing \neq \mathfrak{F} \subseteq \mathfrak{B}$, and let $B = \cap\mathfrak{F}$. Fix $F \in \mathfrak{F}$; int B is an open subset of the resolvable set int F and is therefore resolvable. Plainly B is closed, so that, by Lemma 1, $B \in \mathfrak{B}$.

Now fix $A, B \in \mathfrak{B}$, and let $V = [\operatorname{int}(A \cup B)]\setminus \operatorname{cl}\operatorname{int} A$, an open set. If $\varnothing \neq W \subseteq V$ and if W is open, then certainly $W \not\subseteq A$; but then $W\setminus A$ is a nonempty open subset of B, so that $W \cap \operatorname{int} B \neq \varnothing$. It follows that every nonempty open set $W \subseteq \operatorname{int}(A \cup B)$ meets either int A or int B, so that $\operatorname{int}(A \cup B) \subseteq \operatorname{cl}(\operatorname{int} A \cup \operatorname{int} B)$. Finally, let $D \subseteq \operatorname{int} A$ and $E \subseteq \operatorname{int} B$ be such that D and $(\operatorname{int} A)\setminus D$ (E and $(\operatorname{int} B)\setminus E$, resp.) are dense in $\operatorname{int} A(\operatorname{int} B$, resp.). Let $F = D \cup (E\setminus\operatorname{int} A)$; then F and $(\operatorname{int} A \cup \operatorname{int} B)\setminus F$ are dense in $\operatorname{int} A \cup \operatorname{int} B$, which we have seen to be dense in $\operatorname{int}(A \cup B)$. Thus, $\operatorname{int}(A \cup B)$ is resolvable, so that $A \cup B \in \mathfrak{B}$.

(ii) It suffices to show that $(\mathfrak{T}^\#)^\# \supseteq \mathfrak{T}^\#$, or indeed that $\mathfrak{T}^* \subseteq (\mathfrak{T}^\#)^*$, i.e., that every $\mathfrak{T}$-boundary is a $\mathfrak{T}^\#$ = boundary. Suppose that $B \in \mathfrak{B}$. Then $X\setminus B \in \mathfrak{T}^*$, whence B is $\mathfrak{T}^\#$-closed. Also, $\mathfrak{T}^\# \subseteq \mathfrak{T}$, so that $\operatorname{int}_{\mathfrak{T}^\#}B \subseteq \operatorname{int}_{\mathfrak{T}}B$, and $\operatorname{int}_{\mathfrak{T}^\#}B \in \mathfrak{T}$. Now, $\operatorname{int}_{\mathfrak{T}}B$ is $\mathfrak{T}$-resolvable, so let D and E be complementary $\mathfrak{T}$-dense subsets of $\operatorname{int}_{\mathfrak{T}}B$; clearly $\operatorname{cl}_{\mathfrak{T}^\#}(D \cap \operatorname{int}_{\mathfrak{T}^\#}B) \supseteq \operatorname{cl}_{\mathfrak{T}}(D \cap \operatorname{int}_{\mathfrak{T}^\#}B) \supseteq \operatorname{int}_{\mathfrak{T}^\#}B$, and similarly for E, from which $\operatorname{int}_{\mathfrak{T}^\#}B$ is $\mathfrak{T}^\#$-resolvable, and B is a $\mathfrak{T}^\#$-boundary.

(In connection with (i) we may note that if $B_i = \operatorname{bdry} A_i (i < 2)$ then $B_0 \cup B_1 = \operatorname{bdry} [(A_0 \cap \operatorname{int} B_0\setminus\operatorname{bdry} B_1) \cup (A_1 \cap \operatorname{int} B_1\setminus\operatorname{bdry} B_0) \cup (X\setminus(B_0 \cup B_1))]$ explicitly.) ■

LEMMA 2. *$\mathfrak{T}^\# = \mathfrak{T}$ if and only if every nondense open subset of X is resolvable.*

Proof. Suppose first that $\mathfrak{T}^\# = \mathfrak{T}$, and let V be a nondense open subset of X. Let $W = X\setminus\operatorname{cl} V$; since V is nondense, W is nonempty. By assumption, therefore, $W \in \mathfrak{T}^*$, so that $\operatorname{cl} V \in \mathfrak{B}$. It follows from Lemma 1 that $\operatorname{int}\operatorname{cl} V$ is resolvable. But V is an open subset of $\operatorname{int}\operatorname{cl} V$, whence V is also resolvable.

Conversely, suppose that every nondense open subset of X is resolvable, and let $V \in \mathfrak{T}$. Certainly $V \in \mathfrak{T}^\#$ if $V = \varnothing$, so we assume that $V \neq \varnothing$. But then $X\setminus V$ is a closed set whose interior, being nondense and open, is resolvable. By Lemma 1, therefore, $X\setminus V \in \mathfrak{B}$, so that $V \in \mathfrak{T}^* \subseteq \mathfrak{T}^\#$. Since $\mathfrak{T}^\# \subseteq \mathfrak{T}$ in any case, $\mathfrak{T}^\# = \mathfrak{T}$. ■

THEOREM 2. *$\mathfrak{T}^\# = \mathfrak{T}$ if and only if either X is resolvable, or every nonempty open subset is dense in X.*

Proof. If every nonempty open subset of X is dense in X, then $\varnothing$ is the only nondense open subset of X; and $\varnothing$ is resolvable, so that $\mathfrak{T}^\# = \mathfrak{T}$ by Lemma 2. If, on the other hand, X is resolvable, then every open subset of X is resolvable, and again Lemma 2 implies that $\mathfrak{T}^\# = \mathfrak{T}$.

To prove the converse, assume that $\mathfrak{T}^\# = \mathfrak{T}$, and let V be a nonempty, nondense open subset of X. (If no such V exists, we are done.) Use Zorn's Lemma to find a maximal disjoint collection

$\mathcal{V} \subseteq \mathfrak{T}$ such that $V \in \mathcal{V}$. Each member of $\mathcal{V}$ is clearly nondense and therefore (by Lemma 2) resolvable; and $\mathcal{V}$ is disjoint, whence $\cup\mathcal{V}$ is resolvable as well. Finally, the maximality of $\mathcal{V}$ implies that $\cup\mathcal{V}$ is dense in X, whence X must be resolvable. ■

COROLLARY. *For X a nondegenerate Hausdorff space, $\mathfrak{T}^{\#} = \mathfrak{T}$ if and only if X is resolvable.* ■

Recall that a point $x \in X$ is isolated if $\{x\}$ is an open subset of X. Clearly a resolvable space has no isolated points. Suppose that x is an isolated point of X; if $\mathrm{cl}\{x\} \neq X$, then by Theorem 2 $\mathfrak{T}^{\#} \neq \mathfrak{T}$. In particular, $\mathfrak{T}^{\#} \neq \mathfrak{T}$ whenever either X contains at least two isolated points, or X is a nondegenerate T_1-space with at least one isolated point. However, the Sierpiński space $(X = \{0,1\}, \mathfrak{T} = \{X, \varnothing, \{0\}\})$ is an example of a space for which $\mathfrak{T}^{\#} = \mathfrak{T}$ despite the presence of an isolated point. To avoid such pathologies we shall assume for the remainder of the discussion that X is a nondegenerate T_1-space.[1]

Familiar spaces without isolated points (like the Euclidean spaces, the Hilbert Cube, the Cantor Space, and the spaces of rational and irrational numbers) are all resolvable, so that their topologies are their "boundary topologies." However, there are irresolvable Hausdorff (even T_3) spaces without isolated points [**1**]. (One particularly simple construction begins with the usual topology, $\mathfrak{S}$, on $\mathbf{R}$. Let $\mathbf{T} = \{\mathfrak{T}: \mathfrak{S} \subseteq \mathfrak{T}, \mathfrak{T}$ is a T_3 topology on $\mathbf{R}$ without isolated points$\}$. Check that Zorn's Lemma applies to $\langle \mathbf{T}, \subseteq \rangle$, and pick a maximal $\mathfrak{T} \in \mathbf{T}$. Then $\langle \mathbf{R}, \mathfrak{T} \rangle$ is irresolvable, for if D and $\mathbf{R} \setminus D$ were $\mathfrak{T}$-dense subsets of $\mathbf{R}$, the topology generated by $\mathfrak{T} \cup \{D, \mathbf{R} \setminus D\}$ would belong to $\mathbf{T}$, contradicting the maximality of $\mathfrak{T}$.) The following characterization is known.

THEOREM 3. [**2**] *Let $\langle X, \mathfrak{T} \rangle$ be a space without isolated points. Then X is irresolvable if and only if $\mathfrak{T}$ contains a base, $\mathfrak{B}$, for an ultrafilter on X. (In fact, we may take $\mathfrak{B}$ to be* $\{\mathrm{int}\, D$: *$cl\, D = X$*$\}$.) ■

Theorem 3 can be used to show that the corollary to Theorem 2 cannot be extended to T_1-spaces. For if $\mathfrak{T}$ is a free, or nonprincipal, ultrafilter on $\mathbf{N}$, the set of natural numbers, then $\langle \mathbf{N}, \mathfrak{T} \rangle$ is an irresolvable, nondegenerate T_1-space (by Theorem 3), yet $\mathfrak{T}^{\#} = \mathfrak{T}$ (by Theorem 2).

(A somewhat different picture emerges if in the definition of $\mathfrak{T}^*$ and $\mathfrak{T}^{\#}$ we replace $\mathfrak{B}$ by the set of boundaries of $\mathfrak{T}$-*open* subsets of X. For these are just the $\mathfrak{T}$-(nowhere dense closed) subsets of X, and $\mathfrak{T}^*$ becomes the family of $\mathfrak{T}$-(dense open) subsets of X. Thus, with this new definition of $\mathfrak{T}^{\#}$ Theorem 1 remains valid, but $\mathfrak{T}^{\#} = \mathfrak{T}$ if and only if every nonempty open subset of X is dense.)

References

1. E. Hewitt, A problem of set-theoretic topology, Duke Math. J., 10(1943) 309–333.

2. V. I. Malyhin, Products of ultrafilters, and indecomposable spaces, Mat. Sb. (N.S.), 90(132) (1973) 106–116, 166.

[1]We thank the referee for drawing our attention to this difficulty.

MISCELLANEA

73.

There once was a Lemma of Dehn
Which caused us great trial and pain
For the proof, it would topple us.
Papakyriakopoulos
With a tower at last made things plain.

—FREDERICK NORWOOD, Institute for Advanced Study

ON THE INADEQUACY OF COFINAL SUBNETS AND TRANSFINITE SEQUENCES

E. S. WOLK
Department of Mathematics, University of Connecticut, Storrs, CT 06268

A basic theorem of general topology states that if *p is a cluster point of a net S, then there is a subnet of S that converges to p* [**4**, p. 71], [**7**, p. 75]. In [**4**, p. 77] Kelley gives an example, due to Arens [**1**], to show that in this theorem the net S need not have a *cofinal* subnet that converges to the cluster point p. Another example has been given by Priestley [**5**]. These two known examples probably impress many students as being rather complicated or pathological. We show in this note that a familiar source of counterexamples in topology, namely the "Tychonoff Plank," can be readily adapted to provide a simple example of the above type.

We follow [**4**] and [**7**] in our terminology and notation concerning nets. The reader will recall that if $S = \{S(n), n \in D\}$ is a net defined on the directed set D, and if E is any cofinal subset of D, then the restriction of S to E is called a *cofinal subnet* of S.

Let N denote the nonnegative integers, and W the set of all ordinals less than Ω, the first uncountable ordinal. Let $N^* = N \cup \{\omega\}$, and $W^* = W \cup \{\Omega\}$, each provided with the usual order topology. Consider the space $X = W^* \times N^*$ in its product topology, and let p be the point (Ω, ω) in X. Note that a base for the neighborhood system of p in the product topology consists of all subsets of X of the form

$$U(\beta, n) = \{(\alpha, m) \in X: \alpha > \beta, m > n\},$$

for $\beta \in W$ and $n \in N$.

Now let D denote the set $W \times N$ with its lexicographic ordering: that is, $(\alpha, m) > (\beta, n)$ in D if and only if (i) $\alpha > \beta$, or (ii) $\alpha = \beta$ and $m > n$. Define a net S on D into X by $S(\alpha, m) = (\alpha, m)$ for all (α, m) in D. Since D is a well-ordered set, the net S is a "transfinite sequence." Furthermore, each member of D has only countably many predecessors (and hence D is isomorphic to W). It is clear that p is a cluster point of S, because S is frequently in every set $U(\beta, n)$. However, no cofinal subnet of S converges to p. For suppose that T is any cofinal subnet of S. Then the domain of T is an uncountable subset of D, and hence there exists $k \in N$ such that the set $L_k = \{(\alpha, k): \alpha \in W\}$ contains uncountably many values of T. But then T is frequently in L_k, and so T cannot converge to p.

The space $X = W^* \times N^*$ also provides another useful counterexample in topology. It is of course well known that sequences are "inadequate" for topology, in the sense that the closure of a set A (even if A is countable) cannot always by characterized as the set of all points that are limits of convergent sequences in A. (Counterexamples are given in [**3**], [**4**, p. 77], [**5**], and [**6**].) Birkhoff [**2**, Theorem 10] has shown that *transfinite* sequences are also inadequate for this purpose. Our space X provides still another counterexample of this type; for if A is the subset $W \times N$ of X, then p is an accumulation point of A, but no transfinite sequence in A converges to p.

To prove this assertion, suppose that $S = \{(x(\lambda), y(\lambda)), \lambda \in E\}$ is a transfinite sequence in A that converges to p. Clearly the range of any such transfinite sequence S must be uncountable, and so E cannot have a countable cofinal subset. For each $m \in N$, define $E_m = \{\lambda \in E: y(\lambda) = m\}$. We assert that E_m is cofinal in E for some m. For if not, let $\lambda_m =$ l. u. b. E_m for each $m \in N$. Since $\cup\{E_m: m \in N\} = E$, the countable set $\{\lambda_m: m \in N\}$ must be cofinal in E, which is impossible. So choose an m such that E_m is cofinal in E. Since $\{x(\lambda), \lambda \in E\}$ converges to Ω, so does $\{x(\lambda), \lambda \in E_m\}$; and hence the transfinite sequence $\{(x(\lambda), y(\lambda)), \lambda \in E_m\}$ converges to (Ω, m). This is a contradiction, since this subnet of S must converge to $p = (\Omega, \omega)$.

References

1. R. Arens, Note on convergence in topology, Math. Mag., 23 (1950) 229–234.
2. G. Birkhoff, Moore-Smith convergence in general topology, Ann. of Math., 2, 38 (1937) 39–56.
3. R. O. Davies, Limits of sequences of real functions, Quart. J. Math. Oxford Ser., 2, 23 (1972) 57–62.

4. J. L. Kelley, General Topology, Springer-Verlag, New York, 1955.
5. W. M. Priestley, Nets and sequences, an example, this MONTHLY, 75 (1968) 1098–1099.
6. ______, A sequentially closed countable dense subset of I^I, Proc. Amer. Math. Soc., 24 (1970) 270–271.
7. S. Willard, General Topology, Addison-Wesley, Reading, Mass., 1970.

ON THE LAGRANGE REMAINDER OF THE TAYLOR FORMULA

ALFONSO G. AZPEITIA
Department of Mathematics, University of Massachusetts, Boston, MA 02125

We prove an elementary result on the asymptotic behavior, when the length of the interval involved approaches zero, of the "intermediate point" (or points) whose existence is ensured by the Lagrange-Taylor formula or, in particular, by the mean value theorem. We assume that a and x are contained in some interval and that the derivative $f^{(n+p)}(\ )$, $(n \geqslant 1, p \geqslant 1)$, exists in that interval and is continuous at a. Therefore, we can expand $f(x)$ about the point a up to the nth power by the Lagrange-Taylor formula to obtain

$$f(x) = f(a) + f'(a)(x-a) + \cdots + [1/(n-1)!]f^{(n-1)}(a)(x-a)^{n-1} + (1/n!)f^{(n)}(\gamma)(x-a)^n \tag{1}$$

with γ strictly between a and x. Our result is now as follows.

THEOREM. *If $f^{(n+j)}(a) = 0$ for all $1 \leqslant j < p$ and $f^{(n+p)}(a) \neq 0$, then*

$$\lim_{x \to a} [(\gamma - a)/(x-a)] = [n!p!/(n+p)!]^{1/p} = \binom{n+p}{n}^{-1/p}$$

Proof. Under the given assumptions it is possible to apply the Lagrange-Taylor formula again to expand $f^{(n)}(\gamma)$ in (1) up to the pth power and then once more to expand $f(x)$ up to the $(n+p)$th power. (See, for example, [2] or [1, p. 174, Problem 74].) We thus obtain

$$f(x) = f(a) + f'(a)(x-a) + \cdots + [1/(n-1)!]f^{(n-1)}(a)(x-a)^{n-1} + (1/n!)\left[f^{(n)}(a) + (1/p!)f^{(n+p)}(\gamma_1)(\gamma - a)^p\right](x-a)^n \tag{2}$$

with γ_1 strictly between a and γ, and

$$f(x) = f(a) + f'(a)(x-a) + \cdots + (1/n!)f^{(n)}(a)(x-a)^n + [1/(n+p)!]f^{(n+p)}(\gamma_2)(x-a)^{n+p} \tag{3}$$

with γ_2 strictly between a and x. From (2) and (3) it follows that

$$[1/(n!p!)]f^{(n+p)}(\gamma_1)(\gamma - a)^p = [1/(n+p)!]f^{(n+p)}(\gamma_2)(x-a)^p$$

and therefore that

$$[(\gamma - a)/(x-a)]^p = [n!p!/(n+p)!]f^{(n+p)}(\gamma_2)/f^{(n+p)}(\gamma_1).$$

As $x \to a$ both $\gamma_1 \to a$ and $\gamma_2 \to a$ and the result follows.

REMARKS and EXAMPLES. (a) In particular if $n = p = 1$, $f''(a) \neq 0$ and $f''(x)$ exists and is continuous as required, then $f(x) = f(a) + f'(\gamma)(x-a)$ with γ asymptotically equal to $(x+a)/2$. More explicitly, given any $\varepsilon > 0$ there is a $\delta > 0$ so that

$$|\gamma - (x+a)/2| < \varepsilon |x-a| \quad \text{when} \quad |x-a| < \delta.$$

(b) The condition $f^{(n+p)}(a) \neq 0$ is necessary. For if $n = p = 1$ and $f(x) = cx + d$, then γ can be chosen anywhere in the interval with end points a and x, for example, such that

$(\gamma - a)/(x - a) = 1/3$, and the conclusion of the theorem, which would be in this case $\lim_{x\to a}[(\gamma - a)/(x - a)] = 1/2$, does not hold.

(c) Since for large p and fixed n, $\binom{n+p}{n}^{1/p} \sim 1$, we have the result that for polynomials with large gaps γ is comparatively closer to x.

(d) As a final example let $f(x) = x^N + x^{N+P}$ with $a = 0$, $n = N$, $p = P$. A short calculation shows that in this case we obtain the equality

$$(\gamma - a)/(x - a) = \binom{n+p}{n}^{-1/p}.$$

The author wishes to thank the referees for suggesting several improvements and for the correction of a computational error.

References

1. G. B. Thomas and R. L. Finney, Calculus and Analytic Geometry, Addison-Wesley, Reading, Mass., 1979.
2. J. Wolfe, A proof of Taylor's formula, this MONTHLY, 60 (1953) 415.

MISCELLANEA

74.

I can scarcely bear to write on mathematics or mathematicians. Oh for words to express my abomination of that science, if a name sacred to the useful and embellishing arts may be applied to the perception and recollection of certain properties in numbers and figures! Oh that I had to learn astrology, or demonology, or school divinity; oh that I were to pore over Thomas Aquinas, and to adjust the relation of Entity with the two Predicaments, so that I were exempted from this miserable study! "Discipline" of the mind! Say rather starvation, confinement, torture, annihilation! But it must be. I feel myself becoming a personification of algebra, a living trigonometrical canon, a walking table of logarithms. All my perceptions of elegance and beauty gone, or at least going. By the end of the term my brain will be "as dry as the remainder biscuit after a voyage." Oh to change Cam for Isis! But such is my destiny; and since it is so, be the pursuit contemptible, below contempt, or disgusting beyond abhorrence, I shall aim at no second place. But three years! I can not endure the thought. I can not bear to contemplate what I must have to undergo. Farewell, then, Homer and Sophocles and Cicero.

Farewell, happy fields,
Where joy forever reigns! Hail, horrors, hail,
Infernal world!

How does it proceed? Milton's descriptions have been driven out of my head by such elegant expressions as the following:

$$\text{Cos}\, x = 1 - \frac{x^2}{1\cdot 2} + \frac{x^4}{1\cdot 2\cdot 3\cdot 4} - \frac{x^6}{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}(sic)$$

$$\text{Tan}\,\overline{a+b} = \frac{\text{Tan}\, a + \text{Tan}\, b}{1 - \text{Tan}\, a + \text{Tan}\, b}(sic).$$

—From G. Otto Trevelyan, The Life and Letters of Lord Macaulay, New York, Harper and Brothers, 1875, vol. 1, p. 91. Found by R. T. HOOD, Franklin College, Franklin, IN 46131.

Editorial query: Is math anxiety new?

ON nTH ROOTS OF POSITIVE OPERATORS

NORMAN M. RICE
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

A bounded operator A on a Hilbert space H is positive if $\langle Au, u\rangle \geqslant 0$ for all $u \in H$. In [**1**] an elementary proof (i.e., avoiding the spectral theorem) is given of the following folk theorem:

THEOREM. *Let A be a positive bounded operator on H, and let n be a positive integer. Then there is a positive bounded operator X such that $X^n = A$.*

In this note we give a very short proof of this theorem. Our proof is an adaptation of the old proof of existence of the positive square root of A by Visser [**4**] and employs the modified Newton Method for the algorithm creating the nth root.

First we gather together the basic properties of positive operators that are needed. They all follow by basic arguments in, e.g., [**2**] or [**3**]. Note that property 2 below is actually a consequence of the existence of square roots; consequently, the proof in this note stands when $n \geqslant 3$; when $n = 2$, we may fall back on the original arguments by Visser, which are reproduced in [**2**] and [**3**].

1. The relation defined by "$A \leqslant B$ if and only if $B - A$ is positive" is an order relation on symmetric operators that is compatible with operator addition and scalar multiplication.

2. The order relation $A \leqslant B$ is compatible with multiplication of commuting operators. For example, if $0 \leqslant A \leqslant I$, $0 \leqslant B \leqslant I$, and $AB = BA$, then $0 \leqslant AB \leqslant I$; and, if $A \leqslant B$, $0 \leqslant C$, $AC = CA$, and $BC = CB$, then $CA \leqslant CB$.

3. If A is a positive bounded operator that is not 0, then $(\|A\|)^{-1}A \leqslant I$.

4. If $\{X_k\}$ is a sequence of bounded operators such that $0 \leqslant X_k \leqslant X_{k+1} \leqslant I$, then there is a positive operator X that is the strong limit of $\{X_k\}$, i.e., $\{X_k u\} \to Xu$ for all $u \in H$.

We shall also use twice the following "mean-value" lemma.

LEMMA. *If Y and Z are operators with $0 \leqslant Z \leqslant Y \leqslant I$, and if $YZ = ZY$, then for any positive integer n we have*

$$\frac{1}{n}(Y^n - Z^n) \leqslant Y - Z.$$

Proof of lemma. Since $YZ = ZY$, we may factor $Y^n - Z^n$ to get

$$\begin{aligned} Y^n - Z^n &= (Y - Z)(Y^{n-1} + Y^{n-2}Z + \cdots + Z^{n-1}) \\ &\leqslant (Y - Z)(I + I + \cdots + I) = n(Y - Z). \end{aligned}$$

Proof of theorem. In view of property 3 above, it is sufficient to consider $0 \leqslant A \leqslant I$. Define the sequence $\{X_k\}$ by

$$X_0 = 0, \quad X_{k+1} = X_k + \frac{1}{n}(A - X_k{}^n).$$

We shall see that $\{X_k\}$ is an increasing bounded sequence of positive operators whose limit X satisfies $X^n = A$. To this end we first note that each X_k is a polynomial in A, so that the X_k's all commute with each other. Next we show by induction that the sequence is positive and increasing: certainly $X_1 \geqslant X_0 = 0$, and, assuming $X_{k+1} \geqslant X_k$, we have

$$\begin{aligned} X_{k+2} - X_{k+1} &= X_{k+1} + \frac{1}{n}(A - X_{k+1}^n) - X_k - \frac{1}{n}(A - X_k^n) \\ &= X_{k+1} - X_k - \frac{1}{n}(X_{k+1}^n - X_k^n) \\ &\geqslant 0 \end{aligned}$$

by the lemma. Again by induction we show that $X_k \leqslant I$ for all k: clearly $X_0 \leqslant I$, and, assuming

$X_k \leqslant I$ and applying the lemma again, we have

$$X_{k+1} = X_k + \frac{1}{n}(A - X_k^n) \leqslant X_k + \frac{1}{n}(I - X_k^n) \leqslant X_k + (I - X_k) = I.$$

Consequently $0 \leqslant X_k \leqslant X_{k+1} \leqslant I$ for all k, so that the sequence $\{X_k\}$ has a strong limit X, which clearly satisfies $X = X + \frac{1}{n}(A - X^n)$. Hence $X^n = A$ as required.

References

1. D. R. Brown and M. J. O'Malley, On nth roots of positive operators, this MONTHLY, 87 (1980) 380–382.
2. F. Riesz and B. Sz.-Nagy, Functional Analysis, Ungar, New York, 1955.
3. Martin Schechter, Principles of Functional Analysis, Academic Press, New York, 1971.
4. C. Visser, Note on linear operations, Proc. Acad. Amsterdam, 40 (1937) 270–272.

KANTOROVICH-TYPE INEQUALITIES

ACHIM CLAUSING

Institut für Mathematische Statistik, Universität Münster, 44 Münster, West Germany

Introduction. For a positive random variable Z, having an expectation $E(Z)$, one sometimes approximates the expectation $E(1/Z)$ of its reciprocal by $1/E(Z)$.

The question how to estimate the error in this procedure leads to the inequality of Kantorovich or one of its many variants. (See, for example, **[1]** or **[2]**. The latter article contains a comprehensive bibliography.)

The oldest inequality of this type is due to P. Schweitzer **[7]**:

If $0 < m \leqslant m_i \leqslant M (i = 1, \ldots, n)$, *then*

$$1 \leqslant \left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\right) \leqslant \frac{(m+M)^2}{4mM} = 1 + \frac{(M-m)^2}{4mM} \equiv C_1(m, M). \tag{1}$$

(Strictly speaking, only the right-hand inequality is Schweitzer's inequality. The left-hand sides of (1) and of (2)—(6) below are variants of the arithmetic-harmonic mean inequality; our main interest is in the "complementary" right-hand sides.)

The original inequality of Kantorovich **[5]** appeared more than 30 years later as a means to estimate the rate of convergence of the method of steepest descent for solving systems of linear equations:

If $0 \leqslant \lambda_i \leqslant 1 (i = 1, \ldots, n)$, $\sum_{i=1}^{n} \lambda_i = 1$, *then*

$$1 \leqslant \left(\sum_{i=1}^{n} \lambda_i m_i\right)\left(\sum_{i=1}^{n}\frac{\lambda_i}{m_i}\right) \leqslant C_1(m, M). \tag{2}$$

P. Henrici **[3]** observed that (1) implies (2), and he showed also that equality can hold on the right-hand side of (2) if and only if there is a subset I of $\{1, \ldots, n\}$ such that $\Sigma_{i \in I}\lambda_i = \frac{1}{2}$. Thus neither (1) nor (2) is sharp in general. Inequality (1) and hence also (2) can be derived from another inequality of Schweitzer:

If f is a real function on $(0, 1)$ *such that* $0 < m \leqslant f(x) \leqslant M$, $x \in (0, 1)$, *and f is integrable, then*

$$1 \leqslant \int_0^1 f(x)\,dx \int_0^1 \frac{1}{f(x)}\,dx \leqslant C_1(m, M). \tag{3}$$

[Continued on p. 327.]

C E N T E R   S E C T I O N
(Vol. 89, No. 5, May 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, T(13: 1, 2). College Mathematics for Management, Life, and Social Sciences, Second Edition. Raymond A. Barnett. Dellen Pub, 1981, xiv + 690 pp, $22.95. [ISBN: 0-89517-024-8] Covers topics from algebra, matrices and linear programming, interest, probability and distributions, and calculus (including logs, exponential functions and multivariable optimization). Many exercises use real data, but the depth of application seems very superficial. Solutions manual, computer applications supplement and test battery available. GHM

General, T(12). Mathematics Workbook for College Entrance Examinations, Third Revised Edition. Samuel C. Brownstein. Barron's Educ Ser, 1976, 285 pp, $5.95 (P). [ISBN: 0-8120-0654-2] The third edition is expanded to include practice exercises in the areas of roots and radicals, inequalities, and coordinate geometry. The number of practice aptitude tests has been increased from ten to twenty. JJ

Mathematics Appreciation, T*(14-15: 1), S*, P, L.** Numbers and Infinity: A Historical Account of Mathematical Concepts. Ernst Sondheimer, Alan Rogerson. Cambridge U Pr, 1981, x + 172 pp, $15.95; $7.95 (P). [ISBN: 0-521-24091-3; 0-521-28433-3] A real gem, concise yet conversational, tracing a middle path between the modern, ahistorical approach of typical texts and the comprehensive treatment of historical treatises: an informal account of the evolution of algebra and analysis (the "number" and "infinity" of the title), from classical Greece through, e.g., fractals and nonstandard analysis. Includes lists of term paper topics, and extensive references to related literature. LAS

Precalculus, T(13). College Algebra and Trigonometry. Lawrence G. Gilligan, Robert B. Nenno. Goodyear Pub, 1981, x + 582 pp, $20.95. [ISBN: 0-8302-1992-7] A fairly standard precalculus text with a plethora of worked out sample problems and many exercises. A few calculator and computer exercises are included in many exercise sets. A supplemental student research project list is included with each chapter. PH

Precalculus, T(13: 1), L. Integrated Algebra, Trigonometry, and Analytic Geometry, Fourth Edition. Robert C. Fisher, Allen D. Ziebur. Prentice-Hall, 1982, xii + 500 pp, $21.95. [ISBN: 0-13-468967-4] In this new edition of a relatively long-lived standard text, the authors say "There is no less mathematics,..., we just present it more simply." No major revisions but use is now made of hand-held calculators. (Third Edition, TR, August-September 1972.) JS

Education, P. Results from the Second Mathematics Assessment of the National Assessment of Educational Progress. Ed: Mary Kay Corbitt. NCTM, 1981, v + 167 pp, $12.50 (P). [ISBN: 0-87353-172-8] Integrated presentation of results previously reported piecemeal in NCTM journals. Overview of NAEP goals and procedures and methods to measure change from first assessment. MW

Education, S(16-18). Research Within Reach: Elementary School Mathematics. Mark J. Driscoll. NCTM, 141 pp, $6.25 (P). Introduction to applications of educational research to teaching. Suggestions based on research results with recommendations for further reading. Addresses 21 areas identified as major concerns by elementary mathematics teachers. MW

History, P, L?. Geometry in Ancient and Medieval India. T.A. Sarasvati Amma. Motilal Banarsidass, 1979, xi + 280 pp, Rs. 60. [ISBN: 0-89684-020-4] This doctoral thesis surveys the Samskrt and Prakrt scientific literature of India beginning with the Vedic literature in an attempt to emphasize the geometrical nature of Indian mathematics. JNC

History, L. Speculations on the Fourth Dimension: Selected Writings of Charles H. Hinton. Ed: Rudolf v.B. Rucker. Dover Pub, 1980, xix + 204 pp, $4 (P). [ISBN: 0-486-23916-0] Excerpts from a series of Scientific Romances and other late nineteenth century writings exploring in both fiction and description the notions of tesseract (a term probably coined by Hinton), two-dimensional time, and four-dimensional space. Hinton linked these pre-relativity ideas to then-current speculations about aether and the motion of objects. LAS

History, P, L*. Oeuvres Mathématiques. Georges de Rham. L'Enseignement Math, 1981, 748 pp, 110 Fr. 50 of de Rham's 61 scientific publications, prefaced by a photograph, a brief curriculum vitae, and a list of publications. LAS

History, S, P, L**.** Emmy Noether: A Tribute to Her Life and Work. Eds: James W. Brewer, Martha K. Smith. Pure and Appl. Math., V. 69. Dekker, 1981, x + 180 pp, $14.75. [ISBN: 0-8247-1550-0] A rich, informative collection arranged to commemorate the one hundredth anniversary of Emmy Noether's birth. A fascinating biography by Clark Kimberling fills nearly one-half the volume. This is followed by four personal reflections (by Saunders MacLane, Olga Taussky, B.L. van der Waerden, and P.J. Alexandroff), by five brief Monthly-level surveys of Noether's mathematics, and finally by Noether's address to the 1932 International Congress of Mathematicians. LAS

Foundations, P. Current Issues in Quantum Logic. Ed: Enrico G. Beltrametti, Bas C. van Fraassen. Plenum Pr, 1981, ix + 492 pp, $59.50. [ISBN: 0-306-40652-7] Proceedings of the workshop held at the Ettore Majorana Center in Erice, Sicily, December 2-9, 1979. JAS

Foundations, S(17), P. Vom Mythos der Mathematischen Vernunft. Detlef D. Spalt. Wissenschaftliche Buchgesellschaft, 1981, 396 pp, (P). [ISBN: 3-534-08758-5] A philosophical discussion, in the form of imaginary conversations among four students, of how mathematics develops and what rigor is. Largely illustrated by a study of the history of uniform convergence. Very strongly influenced by the work of Lakatos. JD-B

Foundations, P. Lecture Notes in Mathematics-891: Logic Symposia, Hakone 1979, 1980. Ed: G.H. Müller, G. Takeuti, T. Tugué. Springer-Verlag, 1981, xi + 394 pp, $22 (P). [ISBN: 0-387-11161-1] Research papers mostly in set theory with some proof theory represented. GHM

Combinatorics, P. Combinatorics. Ed: H.N.V. Temperley. London Math. Soc. Lect. Note Series, V. 52. Cambridge U Pr, 1981, 190 pp, $24.50 (P). [ISBN: 0-521-28514-3] Nine invited lectures given at the Eighth British Combinatorial Conference. Topics covered include shift register sequences, connections between designs and codes, geometry of planar graphs. JRG

Combinatorics, P. Lecture Notes in Mathematics-884: Combinatorial Mathematics VIII. Ed: Kevin L. McAvaney. Springer-Verlag, 1981, xiv + 349 pp, $20 (P). [ISBN: 0-387-10883-1] Proceedings of the Eighth Australian Conference on Combinatorial Mathematics held at Deakin University, Geelong, Australia on August 25-29, 1980. JAS

Algebra, P. Lecture Notes in Mathematics-855: Semigroups. Ed: H. Jürgensen, M. Petrich, H.J. Weinert. Springer-Verlag, 1981, v + 221 pp, $14 (P). [ISBN: 0-387-10701-0] Proceedings of a conference held at Oberwolfach, Germany, December 16-21, 1978, consisting of 14 presentations which have not appeared elsewhere. JAS

Algebra, T(16-18: 1, 2), P, L. Introduction to Algebraic K-Theory. John R. Silvester. Math. Series. Chapman and Hall, 1981, xi + 255 pp, $13.95 (P); $29.95. [ISBN: 0-412-23740-7; 0-412-22700-2] A clear and elementary presentation of the topic suitable for following a two-semester course in abstract algebra. The approach is quite detailed but fails to present the rich intuition of the subject. JAS

Algebra, T(16-18). Theory of Categories. Nicolae and Liliani Popescu. Sijthoff & Noordhoff, 1979, x + 337 pp, $35 (P). [ISBN: 90-286-0168-6] A thorough up-to-date treatment of the basics: completions, algebraic and abelian categories. Good references, lots of exercises, and a good index. JAS

Algebra, T(18: 1),S, P. Radicals of Rings. F.A. Szász. Wiley, 1981, 287 pp, $35. [ISBN: 0-471-27583-2] A thorough and largely self-contained survey of the theory of radicals of associative rings, including a number of results obtained since Divinsky's Rings and Radicals first appeared. Contains separate chapters on nil, Jacobson, Brown-McCoy radicals. Extensive bibliography, indexes, and research problems. JS

Finite Mathematics, T(13: 1, 2). Finite Mathematics: A Modeling Approach, Second Edition. J. Conrad Crown, Marvin L. Bittinger. Addison-Wesley, 1981, xiv + 563 pp, $15.95. [ISBN: 0-201-03145-0] New material in the Second Edition includes a chapter on the Mathematics of Finance and an appendix on Programming in Basic. The presentation of the simplex algorithm has been clarified, and the examples and exercises are less "messy." Sections on the Transportation and Assignment Problems have also been added. (First Edition, TR, October 1977; ER, May 1981.) JRG

Finite Mathematics, T(13-14). Finite Mathematics for Management, Life, and Social Sciences, Second Edition. Raymond A. Barnett. Dellen Pub, 1981, xv + 487 pp, $20.95. [ISBN: 0-89517-026-4] Second edition of a text aimed at students with 1 1/2 or 2 years of high school algebra. The applications orientation of the first edition is retained. AWR

Complex Analysis, T(16-18), S, P. Nine Introductions in Complex Analysis. Sanford L. Segal. Math. Stud., V. 53. Elsevier North-Holland, 1981, xvi + 715 pp, $55 (P). [ISBN: 0-444-86226-9] Treats nine topics in complex analysis (e.g., Riemann mapping theorem, Picard theorems, the Bieberbach conjecture) beginning where a one-semester graduate course typically stops. Motivation and historical background for the theorem-proof development are nicely provided by parallel "notes" sections. Appropriate for self-study by well-prepared students. PZ

Complex Analysis, P. Lecture Notes in Mathematics-888: Padé Approximation and its Applications, Amsterdam 1980. Ed: M.G. de Bruin, H. van Rossum. Springer-Verlag, 1981, vi + 383 pp, $20 (P). [ISBN: 0-387-11154-9] Proceedings of a conference held in Amsterdam, The Netherlands on October 29-31, 1980. JAS

Differential Equations, T(18: 1), P. Lecture Notes in Mathematics-840: Geometric Theory of Semilinear Parabolic Equations. Dan Henry. Springer-Verlag, 1981, iv + 348 pp, $22 (P). [ISBN: 0-387-10557-3] Many general results available for ordinary differential equations may be proved for parabolic partial differential equations once the necessary machinery has been constructed. Here it is. And in a very readable, quite accessible form. A nice collection of examples and exercises. Has flair. PH

Differential Equations, T(18: 2), S, P. Isovector Methods for Equations of Balance. Dominic G.B. Edelen. Sijthoff & Noordhoff Inter, 1980, xxi + 513 pp. [ISBN: 90-286-0420-0] Studies classical second order partial differential equations by realization of solutions in appropriate spaces. Extensive appendices make this text essentially self contained. Contains explicit finite computational methods and worked out examples. PH

Differential Equations, T(15-17: 1, 2), S, P, L. Ordinary Differential Equations: Theory and Applications. M. Rama Mohana Rao. Affiliated East-West Pr, 1980, xii + 266 pp, (P). A senior-level text designed as a bridge between theory and application. Many interesting examples, a number of good exercises, but could use more routine exercises. Prerequisites--calculus and linear algebra. Would provide a strong first course in ordinary differential equations. PH

Differential Equations, P. Integral and Functional Differential Equations. Ed: Terry L. Herdman, Harlan W. Stech, Samuel M. Rankin, III. Pure and Appl. Math., V. 67. Dekker, 1981, x + 276 pp, $32.50 (P). [ISBN: 0-8247-1354-0] Papers from a conference held June 18-20, 1979 at West Virginia University at Morgantown. JAS

Differential Equations, P. Singularities in Boundary Value Problems. Ed: H.G. Garnir. D Reidel Pub, 1981, xvi + 377 pp, $49.50. [ISBN: 90-277-1240-9] Proceedings of the NATO advanced study institute held at Maratea, Italy, September 22-October 3, 1980. This volume is intended to provide a unified, detailed, and up-to-date report on recent results in the area. JAS

Differential Equations, P. Elliptic Problem Solvers. Ed: Martin H. Schultz. Academic Pr, 1981, xiii + 444 pp, $27.50. [ISBN: 0-12-632620-7] Invited and contributed papers from a conference held in Santa Fe, New Mexico from June 30-July 2, 1980 dealing with numerical solution of elliptic partial differential equations. JAS

Differential Equations, S(18), P. Extensions of Minimal Transformation Groups. I.U. Bronšteĭn. Sijthoff & Noordhoff, 1979, viii + 319 pp, $47.50. [ISBN: 90-286-0368-9] A study of topological transformation groups and dynamical systems aiming toward applications to non-autonomous differential equations. After an introductory chapter the emphasis is on minimal transformation groups and their extensions. Includes remarks and extensive bibliography. JS

Differential Equations, S(17), P, L. Semidynamical Systems in Infinite Dimensional Spaces. Stephen H. Saperstone. Appl. Math. Sci., V. 37. Springer-Verlag, 1981, 474 pp, $28 (P). [ISBN: 0-387-90643-6] A semidynamical system is a generalization of a dynamical system in that it does not incorporate "past" behavior. The intention in this book is to use the abstract semidynamical system to unify concepts in areas such as non-autonomous differential equations, partial differential equations, and stochastic differential equations. JG

Differential Equations, T(16), S. Differentialgleichungen, 6. Auflage. Lothar Collatz. Teubner Stuttgart, 1981, 287 pp, (P). [ISBN: 3-519-22033-4] A concise, carefully written, modern, introductory text on differential equations, intended chiefly for physicists, engineers and other users of mathematics. JD-B

Numerical Analysis, S(15-17), P. NUMAL: Numerical Procedures in ALGOL 60. Ed. P.W. Hemker. Math Centrum, 1981, [90-6196-217-X]. General Information and Indices. MC Syllabus 47.1, 76 pp, Dfl. 10,50 (P); Volume 1: Elementary Procedures and Volume 2: Algebraic Evaluations. MC Syllabus 47.2, 211 pp, Dfl. 25,20 (P); Volume 3A: Linear Algebra, Part 1. MC Syllabus 47.3, 204 pp, Dfl. 24,15 (P); Volume 3B: Linear Algebra, Part 2. MC Syllabus 47.4, 234 pp, Dfl. 29,40 (P); Volume 4: Analytical Evaluations and Volume 5A: Analytical Problems, Part 1. MC Syllabus 47.5, 200 pp, Dfl. 24,15 (P); Volume 5B: Analytical Problems, Part 2. MC Syllabus 47.6, 199 pp, Dfl. 24,15 (P); Volume 6: Special Functions and Constants and Volume 7: Interpolation and Approximation. MC Syllabus 47.7, 194 pp, Dfl. 24,15 (P). These volumes contain the listings of a comprehensive library of Algol 60 programs for numerical mathematics written at the Mathematical Centre in Amsterdam. Examples illustrating their use have been included as part of the documentation of each of the programs. AO

Functional Analysis, P. Topics in Ergodic Theory. William Parry. Tracts in Math., V. 75. Cambridge U Pr, 1981, x + 110 pp, $23.95. [ISBN: 0-521-22986-3] Intended as a concise and "speedy" introduction to a large number of topics and examples: martingales, information, recurrence, mixing, Bernoulli shifts, entropy, winding numbers, etc. The exposition, though terse, is clear and direct; nevertheless, one is left with the impression that more than 110 pages are required for this task. TAV

Functional Analysis, T(17-18: 1, 2), S, P, L. Functional Analysis. Harro G. Heuser. Trans: John Horváth. Wiley, 1982, xv + 408 pp, $62.95. [ISBN: 0-471-28052-6] Intended as an introduction to functional analysis; emphasis is on classical problems and the development of their solutions. Lengthy treatment of normed linear spaces through the spectral theorem in Hilbert space, followed by several chapters on more general topological vector spaces. Numerous exercises, examples; index, bibliography. JS

Functional Analysis, T(16-17: 2). Applied Functional Analysis. D.H. Griffel. Math. and its Appl. Halsted Pr, 1981, 386 pp, $79.95. [ISBN: 0-470-27196-5] This textbook contains a systematic introduction to the theory of distributions, Green's function techniques, Banach spaces, fixed point theorems, and the theory of operators in Hilbert space. It also includes a brief introduction to the Fréchet calculus and Sobolev spaces. Applications to problems in mechanics, fluid mechanics, etc., are an integral part of the text. Note the price. AO

Algebraic Geometry, P. Real Algebraic Differential Topology, Part 1. Richard S. Palais. Math. Lect. Ser., V. 10. Publish or Perish, 1981, v + 192 pp, $12. [ISBN: 0-914098-19-5] An exploration of the relationship between modern algebraic geometry and differential topology. JG

Algebraic Geometry, S(18), P. The Curves Seminar at Queen's, Volume I. Anthony V. Geramita. Pure and Appl. Math., No. 58. Queen's U, 1981, 140 pp, (P). A collection of four papers, two research and two expository. The research papers: "Seminormality and the Chinese Remainder Theorem" by B. Dayton and "Algebraic curves with ordinary points" by F. Orecchia. The expository papers: "Branches at a singular point of an algebraic curve" by M. Orzech and "Some remarks on multiplicity" by A. Geramita. SG

Algebraic Geometry, T(18), P. Algebraic Geometry: An Introduction to Birational Geometry of Algebraic Varieties. Shigeru Iitaka. Springer-Verlag, 1982, x + 357 pp, $35. [ISBN: 0-387-90546-4] An introductory text which assumes relatively little commutative algebra. The author hopes to provide a more geometric and less algebraic approach to algebraic varieties by emphasizing birational mappings. No exercises. SG

Algebraic Geometry, P. Lecture Notes in Mathematics-887: Non-commutative Algebraic Geometry, An Introduction. Freddy M.J. van Oystaeyen, Alain H.M.J. Verschoren. Springer-Verlag, 1981, vi + 404 pp, $22 (P). [ISBN: 0-387-11153-0] A well-written attempt to lay the foundations for non-commutative algebraic geometry. The first part gives the needed background from non-commutative algebra and sheaf theory. The second part introduces varieties over an algebraic closed field culminating in a non-commutative Riemann-Roch Theorem for curves. SG

Algebraic Geometry, S(17-18), P. Real Elliptic Curves. Norman L. Alling. Math. Studies, V. 54. Elsevier North-Holland, 1981, xi + 349 pp, $36.25 (P). [ISBN: 0-444-86233-1] A captivating partially historical, partially expository, partially research level treatise on elliptic curves and functions defined over the field of real numbers. The author outlines the early history of elliptic integrals and functions, theta functions, discusses work of Euler, Weierstrass and others, and ends with a description of his own work and his work with Greenleaf. SG

Differential Geometry, T(17-18: 1, 2), S*, P*. Global Lorentzian Geometry. John K. Beem, Paul E. Ehrlich. Pure and Appl. Math., V. 67. Dekker, 1981, vi + 460 pp, $45. [ISBN: 0-8247-1369-9] A treatment of Lorentzian geometry for use in theoretical physics. The viewpoint is that of global differential geometry with much attention given to the development of intuition in addition to the usual formal theorems. JAS

Differential Geometry, T(16-18: 1, 2), L. Tensor Analysis on Manifolds. Richard L. Bishop, Samuel I. Goldberg. Dover Pub, 1980, viii + 280 pp, $5.50 (P). [ISBN: 0-486-64039-6] This classical work contains the topology, linear algebra, and multivariable analysis necessary to introduce differential geometry with some physical applications. (Macmillan edition, TR, May 1968; ER, October 1969.) JAS

Geometry, P. Continuous Geometries With a Transition Probability. John von Neumann. Memoirs No. 252. AMS, 1981, vii + 210 pp, $12 (P). A reworking by Israel Halperin of an unfinished 1937 manuscript by von Neumann. von Neumann proposed a system of axioms for a probability-logic system (inspired by quantum mechanics). This version shows that a slight change in these axioms yields a system which can be identified with the set of projection operators in some finite factor of operators in a Hilbert space. JAS

Algebraic Topology, P. Exact Sequences in the Algebraic Theory of Surgery. Andrew Ranicki. Princeton U Pr, 1981, xvii + 863 pp, $16.50 (P). Results in algebraic K-theory and algebraic L-theory leading to new exact sequences and new applications in the area of topological surgery. JAS

Algebraic Topology, P. Multiple Points of Immersed Manifolds. Ralph J. Herbert. Memoirs No. 250. AMS, 1981, xiv + 59 pp, $4.40 (P). An investigation of intersection and self-intersection properties of differentiable maps of manifolds with the goal of finding "certain formulas for the homology classes represented by the self-intersection loci of an immersed submanifold." SG

Algebraic Topology, P. Operations in Connective K-Theory. Richard M. Kane. Memoirs No. 254. AMS, 1981, vi + 102 pp, $6 (P). "The purpose of this paper is to construct a family of operations in complex connective K-theory and to demonstrate their usefulness in analyzing the action of the

Steenrod algebra on the mod p cohomology of certain spectra." JAS

Topology, P. General Topology and Modern Analysis. Ed: L.F. McAuley, M.M. Rao. Academic Pr, 1981, xv + 514 pp, $36. [ISBN: 0-12-481820-X] Proceedings of a conference held in May 1980 at the University of California, Riverside, in honor of the retirement of Professor F. Burton Jones. Includes work on metrization (just before Fleissner's new example), aposyndesis, algebraic and differential topology, and modern analysis and set theory. JAS

Topology, P. Approximation Results in Topological Manifolds. T.A. Chapman. Memoirs No. 251. AMS, 1981, iii + 64 pp, $4 (P). Results concerning the approximation of maps between topological manifolds by homeomorphisms, approximate fibrations, and block bundles. JAS

Operations Research, S(13), L. Elements of Project Management: Plan, Schedule, and Control. M. Spinner. Prentice-Hall, 1981, xi + 212 pp, $19.95. [ISBN: 0-13-269852-8] Presents the basic principles of CPM and PERT planning methods. Illustrated with examples derived from the author's experience as a planning manager for Ford Motor Company. Suggestions for term projects, but few exercises. JRG

Operations Research, P. Applied Operations Research in Fishing. Ed: K. Brian Haley. Plenum Pr, 1981, xvi + 490 pp, $59.50. [ISBN: 0-306-40634-9] Proceedings of NATO Symposium on Applied Operations Research in Fishing, 1979. Contributed papers focus on various models for fishery management and for physical and biological processes in the Barents Sea. JRG

Operations Research, T(14: 1), L. Linear Programming: Algorithms and Applications. S. Vajda. Chapman & Hall, 1981, vii + 150 pp, $9.95 (P). [ISBN: 0-412-16430-2] Thorough discussion of various algorithms, including simplex, revised simplex, and dual simplex. Covers use of LP in game theory and network flow problems, and integer programming generally. Few problems. JRG

Operations Research, T(16-17: 1), S, L. Introduction to Simulation and SLAM. A. Alan B. Pritsker, Claude Dennis Pegden. Halsted Pr, 1979, xviii + 588 pp, $25. [ISBN: 0-470-26588-4] Designed to serve as an introduction to both simulation modeling and the simulation language SLAM. A chapter near the end of the book compares SLAM with several other widely used simulation languages. AO

Probability, T(18: 1), P. Lecture Notes in Mathematics-849: Multiple Wiener-Itô Integrals: With Applications to Limit Theorems. Péter Major. Springer-Verlag, 1981, vii + 127 pp, $9.80 (P). [ISBN: 0-387-10575-1] A study of the multiple Wiener-Itô integrals as a tool for the problem of asymptotic behavior of partial sums of dependent-random variables. Emphasizes the important techniques used in such investigations. A fairly self-contained presentation. PH

Probability, T*(17: 2), L*. A Second Course in Stochastic Processes. Samuel Karlin, Howard M. Taylor. Academic Pr, 1981, xviii + 542 pp, $35. [ISBN: 0-12-398650-8] Assumes background at the level of the authors' First Course (TR, December 1975). This volume contains an incredible amount of material, tightly packed and requiring careful reading. Several applications of each topic are developed. Many new theoretical results appear here for the first time in a text. TAV

Probability, P. Lecture Notes in Mathematics-850: Séminaire de Probabilités XV, 1979/80. Ed: P. Azéma, M. Yor. Springer-Verlag, 1981, iv + 704 pp, $35 (P). [ISBN: 0-387-10689-8] In addition to the 48 presentations from 1979/80, this volume contains listings of the contents of Volumes I-XIV, 1966/67 to 1978/79. JAS

Probability, P. Lecture Notes in Mathematics-876: Ecole d'Eté de Probabilités de Saint-Flour IX-1979. J.P. Bickel, N. El Karoui, M. Yor. Springer-Verlag, 1981, ix + 280 pp, $13.20 (P). [ISBN: 0-387-10860-2] Three invited lectures: "Quelques Aspects de la Statistique Robuste" by J.P. Bickel; "Les Aspects Probabilistes du Contrôle Stochastique" by N. El Karoui; and "Sur la Théorie du Filtrage" by M. Yor. JAS

Statistics, T(15-17: 1, 2), S, L. Intermediate Statistical Methods. G. Barrie Wetherill. Chapman and Hall, 1981, xvi + 390 pp, $25. [ISBN: 0-412-16440-X] The first few chapters review basic statistical ideas. Later ones take up multiple regression, polynomial regression, analysis of variance, components of variance, and crossed classifications. Presupposes calculus and matrix algebra. FLW

Statistics, T(13-15: 1-3), S, L*. Biometry: The Principles and Practice of Statistics in Biological Research, Second Edition. Robert R. Sokal, F. James Rohlf. W.H. Freeman, 1981, xviii + 859 pp, $29.95. [ISBN: 0-7167-1254-7] A non-mathematical, non-theoretical treatment with biological examples of elementary statistics, analysis of variance, analysis of frequencies, and multiple regression. The new edition reflects the availability of calculators and computer packages as well as recent developments in statistics. (First Edition, TR, May 1970.) FLW

Statistics, T(15-16: 1), S, L. Intermediate Mathematical Statistics. G.P. Beaumont. Chapman and Hall, 1980, xvii + 248 pp, $10.95 (P). [ISBN: 0-412-15480-3] Sufficiency and efficiency for estimators, elementary decision theory, Bayesian estimators, uniformly most powerful tests, minimax and Bayesian tests, likelihood ratio tests, interval estimation, and other topics. Presupposes calculus. Uses matrices only in optional sections. FLW

Statistics, S(16-17), P. Exercices Commentés de Mathématiques pour l'Analyse Statistique des Données. Jean-Pierre Nakache, Anne Chevalier, Vincent Morice. Dunod, 1981, xvi + 312 pp, (P). [ISBN: 2-04-011497-1] Multivariate analysis and some of the mathematics (largely linear algebra) needed for it, presented in the form of problems and solutions and intended for those who apply statistics. JD-B

Statistics, T(16-18: 1), S, P, L. Factorial Designs. B.L. Raktoe, A. Hedayat, W.T. Federer. Wiley, 1981, xii + 209 pp, $29.95. [ISBN: 0-471-09040-9] A comprehensive treatment of the modern theory of factoral designs for statisticians or mathematicians familiar with matrix algebra and linear models. FLW

Statistics, T(15-16: 1-2), S, L. Statistical Modeling Techniques. Samuel S. Shapiro, Alan J. Gross. Statistics, V. 38. Dekker, 1981, xii + 367 pp, $29.75. [ISBN: 0-8247-1387-7] A brief introduction to the basic ideas of probability and statistics followed by discussion of several continuous and several discrete distributions; the generalized lambda family, the Johnson system, and the Pearson family of empirical models; and regression tests, distance tests and chi-square tests. A final chapter considers simulation and propagation of moments. Presupposes only calculus. FLW

Statistics, T(13-14: 1, 2). Elementary Business Statistics: The Modern Approach, Fourth Edition. John Freund, Frank J. Williams. Prentice-Hall, 1982, xvii + 606 pp, $23.95. [ISBN: 0-13-253120-8] "Decision making in the face of uncertainty" continues as the focus of this new addition. The only substantial changes are the addition of a section on small-sample tests (nonparametric) and a revision of its previous presentation of topics within the chapter on probability. (Second Edition, TR, August-September 1972; Third Edition, TR, March 1978.) JJ

Statistics, P. Applied Time Series: Analysis II. Ed: David F. Findley. Academic Pr, 1981, xii + 798 pp, $49.50. [ISBN: 0-12-256420-0] Elaborations, in some cases substantial, of addresses given at a symposium sponsored by the University of Tulsa and the Tulsa section of the IEEE which was held in Tulsa, Oklahoma on March 3-5, 1980. JAS

Statistics, T(17-18: 1), S*, P*, L*. Recent Advances in Regression Methods. Hrishikesh D. Vinod, Aman Ullah. Statistics, V. 41. Dekker, 1981, xii + 361 pp, $39.50. [ISBN: 0-8247-1574-8] Integrates classical and Bayesian approaches to estimation. Ridge and Stein estimation. Heteroscidasticity. Canonical correlations, simultaneous equations models. Multicollinearity. FLW

Statistics, T(14-17: 1, 2), S, L*. A Second Course in Business Statistics: Regression Analysis. William Mendenhall, James T. McClave. Dellen Pub, 1981, xiii + 637 pp, $24.95. [ISBN: 0-89517-027-2] Reviews briefly the basic ideas of statistics. Simple and multiple regression. Model building. Residue analysis. Analysis of variance. Several case studies from business. Time series. Optimization. Coding. FLW

Statistics, T(13-14: 1). A First Course in Business Statistics. James T. McClave, P. George Benson. Dellen Pub, 1981, xv + 412 pp, $20.95. [ISBN: 0-89517-022-1] Descriptive statistics, probability, discrete and continuous random variables, sampling, estimation and hypothesis testing for one and two samples. Elements of decision theory. Presents many case studies. Presupposes no college mathematics. FLW

Statistics, T(15-17: 1, 2). Statistical Methods for the Social Sciences. Alan and Barbara Finlay Agresti. Dellen Pub, 1979, xvi + 554 pp, $24.95. [ISBN: 0-89517-014-0] Introductory but sophisticated text covering a wide range of topics. Emphasis is on multivariable methods of use to social scientists. Distinguished from other texts by (1) integration of descriptive and inferential methods throughout, and (2) omission of many traditional computational short-cuts in favor of thorough discussion of how to interpret output from computer packages. GHM

Statistics, T*(16-17: 1, 2), P, L. Statistical Theory and Inference in Research. T.A. Bancroft, Chien-Pai Han. Statistics, V. 39. Dekker, 1981, xiv + 372 pp, $34.50. [ISBN: 0-8247-1400-8] Considerably revised and expanded version of Part I of R.L. Anderson and T.A. Bancroft's 1952 text Statistical Theory in Research. Includes new chapters on sampling and inference based on conditional specification, and chapters on regression analysis and analysis of variance, originally covered in Part II. Although designed for applied statisticians, the mathematical treatment is thorough. Special price for class adoptions. RSK

Statistics, P. Statistical Distributions in Scientific Work. Ed: Charles Taillie, Ganapati P. Patil, Bruno A. Baldessari. Reidel Pub, 1981, $156 set. Volume 4: Models, Structures, and Characterizations, xx + 455 pp [ISBN: 90-277-1332-4]; Volume 5: Inferential Problems and Properties, xxii + 439 pp [ISBN: 90-277-1333-2]; Volume 6: Applications in Physical, Social, and Life Sciences, xxii + 455 pp [ISBN: 90-277-1334-0]. Proceedings of the NATO Advanced Study Institute held at Università degli Studi di Trieste, Trieste, Italy from July 10-August 1, 1980. Issued as a sequel to the three earlier volumes (1975) with the same central title. JAS

Computer Literacy, P, L. The Computerization of Society: A Report to the President of France. Simon Nora, Alain Minc. MIT Pr, 1981, xx + 186 pp, $4.95 (P). [ISBN: 0-262-64020-1] "Télématique" is the term used to represent the growing interconnection between computers and telecommunications. This text, written as a report (1978) to the President of France, prescribes and proposes adoption of "a unified national policy to utilize the new technology of télématique." Contains an interesting analysis of the computer's impact on social structures and political systems. JJ

Computer Literacy, P, L*. A Dictionary of Minicomputing and Microcomputing. Philip E. Burton. Garland STPM Pr, 1982, xvii + 346 pp, $42.50. [ISBN: 0-8240-7263-4] A revised version of the well-received A Dictionary of Microcomputing, supplemented by eight appendix glossaries on such special areas as structured programming and Winchester technology. The new organization reflects cheap editing: instead of a single alphabetical list, there are now nine. Nevertheless, this is an impressive and very useful volume. LAS

Computer Programming, T(13-14: 1), S. Simple PASCAL. James J. McGregor, Alan H. Watt. Computer Sci Pr, 1981, viii + 182 pp, $10.95 (P). [ISBN: 0-914894-72-2] A genuinely introductory text, avoiding all computer detail that is not absolutely essential to Pascal. Covers loops, arrays, conditionals and procedures; uses a variety of examples and exercises, none of them mathematical. LAS

Computer Programming, S(13-18), P, L*. 68000 Assembly Language Programming. Gerry Kane, Doug Hawkins, Lance Leventhal. Osborne/McGraw-Hill, 1981, xi + 588 pp, $16.99 (P). [ISBN: 0-931988-62-4] One more well done and very useful manual from the Leventhal, et. al., Osborne/McGraw-Hill formula for books on assembly language programming of standard microcomputers. The extensive indexing and tabulation, and clear but brief general discussion of microcomputer features in this volume makes it an ideal source for getting at the heart (programmatically) of the 68000. JAS

Computer Programming, S*(13-14), L. CBASIC: User Guide. Adam Osborne, Gordon Eubanks, Jr., Martin McNiff. Osborne/McGraw-Hill, 1981, viii + 215 pp, $15 (P). [ISBN: 0-931988-61-6] A nice mixture of tutorial and language reference. Not only is the language itself presented but also its philosophy and ways of interacting with real machines. The "CBASIC is the greatest" theme is a bit overdone. JAS

Computer Programming, L?. Basic Scientific Subroutines, Volume 1. F.R. Ruckdeschel. Byte/McGraw-Hill, 1981, xi + 316 pp, $19.95. [ISBN: 0-07-054201-5] A collection of subroutines written in Basic which implement elementary numerical techniques. Includes subroutines for plotting data on a terminal, manipulating complex variables, performing vector and matrix operations, random variate generation, and series approximations of the elementary functions. A significant shortcoming of this work is the omission of any discussion of the numerical accuracy of the results obtained using these subroutines. AO

Computer Programming, S(13), L. Computer Applications in Finite Mathematics and Calculus. Michael R. Ziegler. Dellen Pub, 1981, 272 pp, $7.95 (P). [ISBN: 0-89517-031-0] To be used in conjunction with an introductory calculus or finite mathematics text; includes an introduction to Basic and 32 Basic programs with example problems and sample runs for each. Exercises require modifications of these or similar new programs. JNC

Computer Programming, S?, L?. Beginner's Guide for the UCSD Pascal System, Third Printing. Kenneth L. Bowles. Byte Books, 1980, ii + 204 pp, $11.95 (P). [ISBN: 0-07-006745-7] Intended as an introduction and reference manual. The approach makes the language seem complicated and intricate. Because UCSD Pascal includes its own special operating system, there is much space devoted to special editors and system commands. There also appears to be no available summary of the language syntax, although there is an appendix comparing UCSD Pascal to standard Pascal. JAS

Computer Science, P. VLSI: Systems and Computations. Ed: H.T. Kung, Bob Sproull, Guy Steele. Computer Sci Pr, 1981, xi + 415 pp, $29.95. Proceedings of a conference held at Carnegie-Mellon University in October, 1981. The six invited and 38 contributed papers cover a wide range of topics in the theory and design of computational systems using VLSI (very large scale integration) technology. AO

Computer Science, S. Proceedings of the Inaugural Conference of the National Computer Graphics Association, Washington, D.C., June 1980. NCGA (2033 M St., NW, Suite 330, Washington, D.C. 20036), 1980, 324 pp, (P). Lots of "show-and-tell" and some generalities about the past, present, and future. No algorithms or circuit diagrams. JAS

Computer Science, S(13-18), L?. Tutorial: Computer Graphics. Kellogg S. Booth. IEEE Computer Soc, 1979, v + 433 pp, $28 (P). A collection of papers on computer graphics from such sources as Scientific American, Computing, Communications of the ACM, and various IEEE journals. Most are expository and suitable for rather general audiences. One "paper" is the third chapter from a popular textbook; clearly coherence played second fiddle to completeness in this compendium. However, the inclusion of massive bibliographies makes this volume useful as a set of pointers to the subject. JAS

Systems Theory, P. Ergodic Theory and Dynamical Systems I: Proceedings Special Year, Maryland 1979-80. Ed: A. Katok. Prog. in Math., V. 10. Birkhäuser Boston, 1981, xi + 333 pp, $18. [ISBN: 3-7643-3036-8] Seven papers by participants in the Special Year. JAS

Applications, P. Stochastic Nonlinear Systems in Physics, Chemistry, and Biology. Ed: L. Arnold, R. Lefever. Series in Synergetics, V. 8. Springer-Verlag, 1981, viii + 237 pp, $29.50. [ISBN: 0-387-10713-4] Proceedings of the workshop held at Bielefeld, Federal Republic of Germany, October 5-11, 1980. Mostly descriptive and qualitative rather than theoretical results. JAS

Applications (Artificial Intelligence), T(17: 1), P. Robot Manipulators: Mathematics, Programming, and Control: The Computer Control of Robot Manipulators. Richard P. Paul. MIT Pr, 1981, 279 pp,

$25. [ISBN: 0-261-16082-X] Brings together, for the first time in uniform notation, selected theories from computer graphics, kinematics, dynamics, statics, control, and programming to provide a working and coherent approach to the overall problem of computer control of mechanical manipulators. GHM

Applications (Biology), S(17-18), P, L. Genealogical and Genetic Structure. C. Cannings, E.A. Thompson. Cambridge U Pr, 1981, xi + 156 pp, $37.50; $14.95 (P). [ISBN: 0-521-23946-X; 0-521-28363-9] A probabilistic model of the effect of genealogical relationships (e.g., patterns of inbreeding) on the genetic characteristics within populations. Discusses such things as genealogical metrics (used to measure distances between population), algorithms for inbreeding coefficients, and strategies for computer simulation. LAS

Applications (Economics), P. A Mathematical Theory of Pure Exchange Economies Without the No-Critical-Point Hypothesis. J.H. van Geldrop. Math. Centre Tracts, V. 140. Math Centrum, 1981, v + 108 pp, Dfl. 13,65 (P). [ISBN: 90-6196-219-6] Results concerning the general structure of the set of first order equilibria and of the set of first order critical Pareto points in a pure exchange economy. Self-contained in the sense that it contains the necessary basics on manifolds and differential topology. JAS

Applications (Engineering), P. Nonlinear Electromagnetics. Ed: Piergiorgio L.E. Uslenghi. Academic Pr, 1980, x + 426 pp, $30. [ISBN: 0-12-709660-4] Solicited papers together with expanded versions of papers from a symposium held as part of the IEEE/URSI International Symposium at the University of Washington, Seattle, in June 1979. JAS

Applications (Engineering), T(16-17: 1), S. Specialist Techniques in Engineering Mathematics. A.C. Bajpai, L.R. Mustoe, D. Walker. Wiley, 1980, ix + 401 pp, $59. [ISBN: 0-471-27907-2] This text provides brief introductions to a number of mathematical subjects of interest to engineers. Among the topics covered are linear systems, stability of systems, optimal control, random processes, Cartesian tensors, the finite element method, the design of experiments, and functional analysis. AO

Applications (Engineering), P. Transactions of the Twenty-Seventh Conference of Army Mathematicians. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC 27709), 1981, xv + 765 pp, (P). 40 papers on the mathematical analysis of combustion and explosive dynamics, from a June 1981 conference at West Point. LAS

Applications (Fluid Mechanics), P. Annual Review of Fluid Mechanics, Volume 14. Eds: Milton van Dyke, J.V. Wehausen, John L. Lumley. Annual Reviews, 1982, 456 pp, $22. [ISBN: 0-8243-0714-3] Sixteen survey articles on such topics as the dynamics of glacial flow, the fluid dynamics of heart valves and the strange attractor theory of turbulence. Opens with a professional biograpical sketch of the Norwegian meteorologist Vilhelm Bjerknes and his students. LAS

Applications (Physics), T(17-18: 1, 2), S, L. Electrodynamics and Classical Theory of Fields & Particles. A.O. Barut. Dover, 1980, xv + 235 pp, $4.50 (P). [ISBN: 0-486-64038-8] A corrected reprint of the 1964 original. Provides a survey of the classical foundations of modern quantum field theory. JAS

Applications (Physics), S(17-18), P. Semi-Classical Approximation in Quantum Mechanics. V.P. Maslov, M.V. Fedoriuk. Trans: J. Niederle, J. Tolar. Math. Physics and Appl. Math., V. 7. D Reidel Pub, 1981, ix + 301 pp, $66. [ISBN: 90-277-1219-0] An exposition of the canonical operator method and some of its applications in the development of multi-dimensional semi-classical approximations. Considers both non-relativistic and relativistic problems. AO

Applications (Physics), T(17-18: 1, 2), P. Non-Relativistic Quantum Dynamics. W.O. Amrein. D Reidel Pub, 1981, viii + 237 pp, $34.50 (P). [ISBN: 90-277-1324-3] Based on a course taught at the University of Geneva to advanced undergraduates in mathematical physics, this gives a fairly self-contained though very sophisticated treatment of time-dependent proofs of a number of results in spectral and scattering theory for Schrödinger operators. The first three chapters contain the basic results on Hilbert spaces, operators, and evolution groups. JAS

Applications (Physics), P. Multiple Scattering and Waves in Random Media. Ed: P.L. Chow, W.E. Kohler, G.C. Papanicolaou. Elsevier North-Holland, 1981, x + 286 pp, $38.25. [ISBN: 0-444-86280-3] Proceedings of a U.S. Army workshop held in Blacksburg, Virginia on March 24-26, 1980. JAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

The Mathematical Association of America

The Sixty Fifth Annual Meeting of the Association

Cincinnati, Ohio

The Sixty Fifth Annual Meeting was held in Cincinnati during the period January 14-17, 1982. There were 2345 registrants including 1298 members of the Association. The meeting was held in conjunction with meetings of the American Mathematical Society and the Association for Women in Mathematics. Sessions of the Association were held in the Convention Center. The Program Committee consisted of Andrew Sterrett, Jr., Chairman, William Beyer, Jessie A. Engle, J. Douglas Faires, Donald O. Koehler, Alan Lazer, Joan P. Leitzel, James H. Wells and Marion Wetzel. The portion of the Program for which abstracts were submitted comprised the following presentations:

Video Tapes for Review and Supplementary Instruction, by John B. Monroe and Cynthia P. Yang.

A project under the sponsorship of the Ohio Mathematical Association of Two-Year Colleges to make 20-minute video cassettes for review and supplementary instruction in mathematics has been funded by the NSF through its CAUSE program. The complete set consists of almost 140 tapes on topics usually taught in two-year colleges, ranging from pre-algebra to calculus, including a few on matrices and a few on differential equations. A unique feature of the tapes is the use of the output of a PLATO computer as an electronic chalk board. Their effectiveness as teaching tools is enhanced by the capability of PLATO to generate graphs and a variety of special effects for directing the viewer's attention.

Report on the 1980 CBMS Survey, Part I: Four-Year Institutions, by Wendell H. Fleming.

Highlights of the 1980 CBMS Survey of undergraduate mathematics in universities and four-year colleges were presented. This is the latest in a series of similar surveys conducted at 5 year intervals. Topics included: growth in mathematical science course enrollments; the rapid growth of computer science; upper division courses; instructional formats; faculty loads; full-time vs. part-time faculty; demographic characteristics of the national mathematical sciences faculty; and administrative organization of departments.

Report on 1980 CBMS Survey, Part II: Two-Year Colleges Under Seige, by Donald J. Albers.

The 1980 CBMS survey of mathematics programs in two-year colleges reveals striking trends: enrollments are up, with computer science accounting for about half the gain; full-time faculty has decreased in size and nearly half are teaching overloads; part-time faculty has doubled in size and exceeds the number of full-time faculty.

Iterated Maps of the Interval, by John Milnor.

Let f be a smooth mapping from a closed interval of real numbers to itself. Typical questions to ask about f are the following. How many periodic orbits does f have? (A periodic orbit is a finite set whose points are cyclically permuted by f.) Does f have periodic orbits A which are stable, in the sense that the successive images under f of any point which is sufficiently close to A must converge towards A? What can one say about the behavior of the sequence x(0), x(1),...of successive images, starting with some random point x(0); where x(i+1) = f(x(i))? The lecture described some of what is known about such questions.

What Do I Know: Current States of One's Knowledge in the Mathematical Fields, by Philip J. Davis.

I am confronted with a mathematical problem P. I think about P and work on it a bit. What then is the state of my knowledge about P and its solution? Here more than thirty "knowledge states" are delineated and studied. They range from "I know how to solve P" to "I can prove that P has no answer" to "I have an algorithm for finding P which seems to work, but I can't prove it" to "I can prove that there are some mathematical 'things' that I cannot compute."

Luzitania: The Birth of Soviet Topology, by Douglas E. Cameron.

The revolutionary atmosphere in Russia during the first two decades of this century was reflected in the classrooms of Moscow University in an exciting new manner of teaching mathematics and brought together an extraordinary group of young mathematicians who were facetiously referred to as "Luzitania." This talk discussed the personalities and motivations of Nikolai Nikolaevich Luzin, founder of the Moscow school, his teacher, Dmitrii Fedorovich Egorov, and two of his famous students, Paul Sergeevich Aleksandrov and Paul Samuelovich Urysohn, who founded the Moscow school of topology.

Computing in an Industrial Research Setting, by Phyllis A. Fox.

The rapid changes which have taken place in computing within an industrial research setting over the last ten years were sketched. This was followed by a description of the current situation--how physical problems arise, are formulated mathematically and then numerically, leading to machine solution. Illustrative examples combining numerical analysis, vector computers and mathematical software libraries were cited.

Putting Calculus in Its Place--A New Curriculum for the First Two Years of College Mathematics, by Anthony Ralston.

The current, fairly standard, two-year calculus-linear algebra sequence is, it was argued, inappropriate for computer science majors and (for the portion they take) for social, behavioral and management science students. For physical science and engineering students, the sequence is far from optimal and, for mathematics majors, it no longer is the best preparation for further study or for the professional careers on which most will embark. The major flaw in the current curriculum is failure to give discrete mathematics an appropriate place. As a step toward rectifying this situation, the talk described the issues to be considered at a conference next summer which will consider the contents of a new curriculum for the first two years of college mathematics.

Math Service Courses Can Be Exciting, by Ronald M. Davis.

The speaker addressed studying and teaching mathematics service courses, courses which support other disciplines. The importance of frequent communication with faculty of the supported program and of a commitment by mathematics faculty to serving the supported area was emphasized. Various approaches to teaching mathematics service courses which increase student retention and the likelihood of additional enrollments were discussed.

Noncomputability in Models of Physical Phenomena--Computable Differential Equations With No Computable Solution, by Marian Boykan Pour-El.

Consider the following question. Can computable data be transformed by a physical process into data which are no longer computable? This question can be made precise. Physical processes suggests differential equations. The concept of computability--e.g., "computable function of a real variable"--has been defined by mathematical logicians.

It can be shown that the above question has an affirmative answer for certain well-known physical processes involving propagation of waves. However, the discussion was much broader. It focused on the interplay among logic, analysis and physical theory. Ideas from these fields lead to the formulation and solution of a variety of problems.

Session in Honor of Martin Gardner, Doris Schattschneider, presider.

Perhaps more than any mathematician could be, Martin Gardner, a nonmathematician, has been a witty and eloquent reporter, interpreter, challenger, correspondent, and prophet for mathematics. For the last twenty-five years, readers have turned to his Scientific American column "Mathematical Games" to enjoy clear, entertaining exposition of a wide range of mathematical ideas--from timeless puzzles to a latest research result. Amateurs and students have been challenged and encouraged by his announcements of readers' newly found examples (or counterexamples) which establish (or topple) ideas. Not only amateurs are thrilled to see their names in his column--more than one mathematician cherishes a one-paragraph mention of his work by Gardner.

Gardner not only has provided stimulation through his published writings but has been the hub of an unheralded service to the mathematical community. His mathematical grapevine (MGG) is his unofficial communication network. Thousands of readers' letters are answered; budding mathematicians are always encouraged. MGG not only gathers and checks information for accuracy, but brings together those working on similar problems. Through this informal service, he has fostered friendships and collaboration between mathematicians who might never otherwise have known of one another.

Algebraic Theory of Penrose's Non-Periodic Tilings of the Plane, by N.G. de Bruijn.

In Scientific American (January 1977) Martin Gardner wrote a fascinating survey on R. Penrose's non-periodic tilings of the plane. Some properties of these tilings remind one of the non-periodic behavior of $\{(n+1)b\} - \{nb\}$ for a fixed irrational b (here $\{nb\}$ denotes the integral part of nb). The speaker showed (Proceedings Kon. Nedrl. Akad. Wetensch. (A) 84 = Indagationes Mathematicae 43, 39-66, 1981) that the shape of a Penrose pattern is determined in a similar way by a complex constant, and properties of the pattern can be derived from properties of that constant.

My Life Among the Polyominoes, by David A. Klarner.

This talk was based on a chapter bearing the same title of "The Mathematical Gardner," a collection of papers written in honor of Martin Gardner. Topics covered were the question of whether there exists a method for finding the number of n-ominoes; the problem of estimating the number of n-ominoes; polyomino jigsaw puzzles; filling boxes with bricks.

Also included on the program were panel discussions entitled "Articulation Efforts Between Universities and Schools" and "NSF Research Institutes: Current Status and Expectations." Other invited speakers and their titles were: "How Computer Animation Changes My Views of Teaching and Research" by Thomas F. Banchoff; "A Tale of Lanczos: Large and Inexact Matrix Computations" by Jane K. Cullum; "Historical Notes of the Calculus" by V. Frederick Rickey; "Recent Solutions of Problems Posed by J. E. Littlewood and N. N. Lusin" by O. Carruth McGehee; "The Game of Dots and Boxes" by E. R. Berlekamp. Past-President Dorothy L. Bernstein presented her Retiring Presidential Address entitled "Mathematical Expectations."

The Association sponsored Mini-Courses entitled "Mathematical Models in Political Science" and "The Use of Computers to Teach Mathematics." There was also an open meeting on secondary school lectureship programs and an open meeting of MAA Section officers.

The Local Arrangements Committee consisted of Frank T. Birtel (ex-officio), S. Elwood Bohn, Thomas J. Bruggeman, Milton D. Cox, Robert M. Dieffenbach, Richard G. Laatsch, William J. Larkin, III, William J. LeVeque (ex-officio), Maita Levine (Publicity Director), H. David Lipsich, Edward P. Merkes (Chairman), Raymond H. Rolwing, and David P. Roselle (ex-officio).

Board and Business Meetings

The Board of Governors met at 9:00 A.M. on Thursday, January 14 in Bronze A of Stouffer's with 40 members present. The major items of business transacted at that meeting will be announced to the membership in the March 1982 issue of FOCUS.

The Business Meeting of the Association was held at 10:00 A.M. on Sunday, January 17 in the North Meeting Room of the Cincinnati Convention Center. The major item of business was presentation of the Award for Distinguished Service to Dr. Thornton Fry. The citation for Dr. Fry was read by Professor G. Baley Price and the complete text has been published in the Monthly.

Election of Members

At its meeting on January 14, 1982, the Board of Governors elected to membership a total of 1063 applicants for individual membership, 4 applicants for academic membership, and 9 applicants for corporate membership. The newly-elected academic and corporate members follow:

Cleveland State University
Felician College
Gettysburg College
Transylvania University

Academic Press, Inc.
Harcourt Brace Jovanovich
John C. Wiley & Sons, Inc.
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Prindle, Weber & Schmidt
Scott, Foresman and Company
Springer-Verlag, Inc.

Respectfully submitted,

David P. Roselle
Secretary

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Louisiana-Mississippi Section

The 1982 meeting of the Louisiana-Mississippi Section was held at the University of Southwestern Louisiana, Lafayette, Louisiana on February 12 and 13, 1982.

Invited Addresses:

"Optional Strategies in Sports," by Leonard Gillman, Univ. of Texas.
"On Choosing a Wife," by Leonard Gillman, Univ. of Texas.

Short Course:

"Graph Theory," by David Berman, Univ. of New Orleans.

Student Papers:

Undergraduate winners of a years's student MAA membership are designated by a plus (+); winners of $50 award plus a year's MAA membership are designated by a double-plus (++).

"A Mathematical Model of an Epidemic," by Karen d'Aquin, Nicholls State Univ..
"Some Interesting Properties of the Taxi-cab Metric," by Julie Bowers, Mississippi Univ. for Women.
+ "The Automation of a Coke Machine Seen as a Finite-State Transducer," by Susan Schof, Loyola Univ..
+ "Classification Under Restrictions," by Loretta Hart, Univ. of New Orleans.

"Estimating the Shapley Value of a Weighted Majority Game by Computer Simulation," by Melanie Rogers, Nicholls State Univ..
"Applications of Group Theory to Molecular Symmetry," by Annette LaRussa, Mississippi Univ. for Women.
"Least Squares Clustering," by Miriam Reilman, Univ. of New Orleans.
"Telegraphic Code as an Injective Homomorphism of Free Monoids," by Heloise B. Olivier, Loyola Univ..
"Some Geometrical Constructions of Cubic Curves," Mary Margaret Freibert, Univ. of Southern Mississippi.
+ "The Brachistochrone Problem," by John M. Graves, Univ. of Southern Mississippi.
++"Solution of a Certain Type of Difference Equation," by Kathleen D. Lopez, Univ. of Southwestern Louisiana.
"Solution of the Schrodinger Wave Equation Representing a Free Particle Using LaPlace Transforms," by Dana Lynn Fields, Mississippi Univ. for Women.
"Remarks on Solving the Differential Equation y" + A Sin y = 0," by Elizabeth A. O'Neal, Univ. of Southern Mississippi.
++"Algebraic Properties of Various Linear Differential Operators," by Altha Elizabeth Blanchet, Univ. of Southwestern Louisiana.

Short Presentations:

* "New Quadrature Formula Designed for Use in Adaptive Algorithms," by Marguerite Saacks, Loyola Univ..
* "The Effect of Unevenly Distributed Catches on Stock Size Estimates Using Cohort Analysis," by S.E. Sims, Louisiana Tech Univ..
"A Pursuit Problem," by Donald M. Bardwell, Nicholls State Univ..
"Error Propogation Analysis of Combustor Data," by Billy J. Holmes, Nicholls State Univ..
* "Best Approximation in Metric Space," by Hideaki Kaneko, Louisiana Tech Univ..
* "The Direct Integral and Poplar Decompositions of Closed Operators," by Clark P. Rhoades, Loyola Univ..
"Examples of Functions of Bounded Variation," by Temple H. Fay and Porter G. Webster, Univ. of Southern Mississippi.
"On a Locally Convex Space of Sequences," by Lalitha Swetharanyam, McNeese State Univ..
"Multisets as Models," by Ed Oxford, Univ. of Southern Mississippi.
* "Fractals," by Margaret M. LaSalle, Univ. of Southern Louisiana.
"A Bivariate Laplace Distribution," by Badiollah R. Asrabadi, Nicholls State Univ..
"A Study of Some Quadratic False Position Root Finding Techniques," by W.W. Watson, Louisiana Tech Univ..
"Mathematics Learning Centers," by Charles D. Miller, American River College.
"Invariant Subspaces for Essentially Unitary Operators," by Ridgley Lange, Univ. of New Orleans.
"Discrete Generalized Cesaro Operators," by Crawford Rhaly, Univ. of Mississippi.
"Lattices of Periodic Functions," by James Caveny, Univ. of Southern Mississippi.
"Characterizing Nowhere Dense Subsets of Perfectly Normal Baire Spaces," by Travis Thompson, Louisiana State Univ. at Eunice.
"A Necessary and Sufficient Condition for the Oscillation of the Bounded Solutions of Higher Order Functional Differential Equations," by J.R. Graef and P.W. Spikes, Mississippi State Univ..
"Linear Difference Equations with Linear Operator Coefficients," by Henry Heatherly, Univ. of Southwestern Louisiana.
* "Discrete Mathematical Structures in a Computer Science Curriculum," by Antonio M. Lopez, Jr., Loyola Univ..
* "The Only Good Function...," by Virginia A. Cook, Nicholls State Univ..
"Commutative Algebras Generated by Idempotents," by David Choate, Univ. of Southwestern Louisiana.
* "Representable Preradicals in Abelian Groups," by Temple H. Fay, Univ. of Southern Mississippi.
* "A Class of Finite Groups with Normality Conditions on Certain Subgroups," by Gary L. Walls, Univ. of Southern Mississippi.
* "Obtaining Mutually Orthogonal Latin Squares from Right Neofields of Characteristic 2," by Corlis Powell Johnson, Jackson State Univ..

Panel Discussions:

"Funded Educational Programs in Mathematics," by Michael Duffy, Univ. of New Orleans, and Steve Doblin, University of Southern Mississippi.
"State-level Developments in Remedial Programs," by Professor A.J. Hulin, Univ. of New Orleans, and Professor John L. Tilley, Mississippi State Univ..
"Classroom Use of Microcomputers." Presiding: David Cook, Univ. of Mississippi, and James Abbott, Univ. of New Orleans. Other participants: Howard Jones, Univ. of New Orleans, and Terry Flaherty, Loyola University.

Special features included a presentation of a plaque to President R.D. Anderson for outstanding contributions made to mathematics (the introduction being made by Professor Paul Rees, a 54-year member of the Association), the showing of the MAWIS video tapes, and a "Cajun Party."

where

$$m_2 = \frac{2m}{M-m} \quad \text{and} \quad M_2 = \left(1 - \frac{m \cdot \ln(M/m)}{M-m}\right)^{-1}.$$

One can verify that $C_4(m, M) > C_2(m, M)$. (Use $C_4(m, M) = C_2(M, m)$.)

It is apparent that, by varying K, μ and ν, one can prove further Kantorovich-type inequalities in the same way.

Since the quasiconvexity of ϕ was the main point of this note, we include the book [6] in our list of references. It contains a short survey of quasiconvex functions.

It remains to give the postponed proof that the set $K = \{f\colon (0,1) \to [m, M]\colon f \text{ concave}\}$ is metrizable and compact with respect to the compact-open topology.

Let f be any concave nonnegative function on (0, 1). Then

$$\frac{-f(y)}{1-y} \leqslant \frac{f(x)-f(y)}{x-y} \leqslant \frac{f(y)}{y}$$

holds for $x, y \in (0,1)$, $x \neq y$. (Interpret the expressions as slopes.) It follows that for all $f \in K$ and $x, y \in (0,1)$

$$|f(x) - f(y)| \leqslant M \cdot (\min(y, 1-y))^{-1} \cdot |x - y|$$

holds, that is, K is equicontinuous. In particular, K is contained in $C((0,1),\mathbb{R})$; and the compact-open topology on this space is metrizable [4, p. 256]. Since K is also closed [6, p. 17] and $\{f(x)\colon f \in K\} = [m, M]$ is compact for all $x \in (0,1)$, we conclude, using the theorem of Arzelà-Ascoli [4, p. 252], that K is compact.

This note was written while the author was a fellow of the Max Kade Foundation at the University of Washington, Seattle.

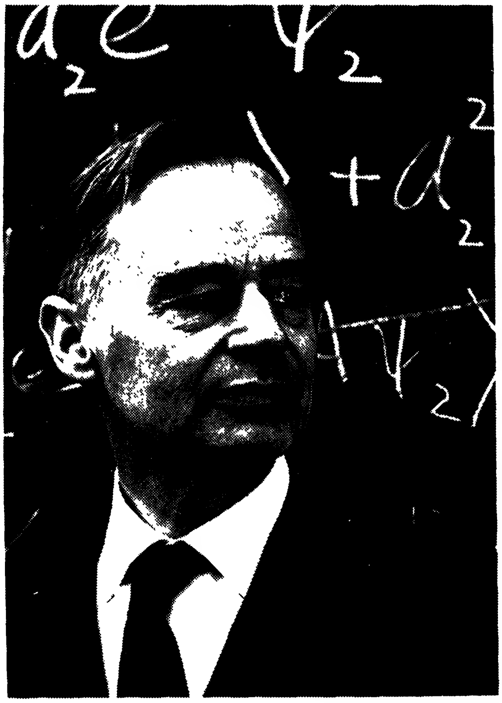
References

1. E. F. Beckenbach, On the inequality of Kantorovich, this MONTHLY, 71 (1964) 606–619.
2. G. T. Cargo, An elementary, unified treatment of complementary inequalities. Inequalities-III, O. Shisha, editor, Academic Press, New York and London, 1972, pp. 39–63.
3. P. Henrici, Two remarks on the Kantorovich inequality, this MONTHLY, 68 (1961) 904–906.
4. T. Husain, Topology and Maps, Plenum Press, New York and London, 1977.
5. L. V. Kantorovich, Functional analysis and applied mathematics (in Russian), Uspehi Mat. Nauk., 3 (1948) 89–185.
6. A. W. Roberts and D. E. Varberg, Convex Functions, Academic Press, New York and London, 1973.
7. P. Schweitzer, An inequality concerning the arithmetic mean (in Hungarian), Math. Phys. Lapok, 23 (1914) 257–261.
8. J. E. Wilkins, The average of the reciprocal of a function, Proc. Amer. Math. Soc., 6 (1955) 806–815.

MISCELLANEA

75. Another feature of this book is the large number of figures. Almost everybody who is looking for a proof draws diagrams that are often incomplete, sometimes inaccurate, but always helpful. Although the French cheerfully consider such diagrams to be incomprehensible to anyone except their authors, the British and Americans often publish them because they think, and I believe they are right, that what has served as scaffolding for an argument may help in understanding it.

—H. LEBESGUE, reviewing *The Theory of Sets of Points*, by W. H. and G. C. Young, in *Bull. Sci. Math.*, (2) 21 (1907) 129.



Two aspects of algebra. (See p. 340.)

References

1. G. Lejeune Dirichlet, *Über die Bestimmung der mittleren Werthe in der Zahlentheorie*, Abhandlungen der Akademie der Wissenschaften zu Berlin, 1849, pp. 69-83. [G. Lejeune Dirichlet's Werke, Band II, Berlin, 1897, pp. 49-66.]

2. É. Lucas, *Sur l'emploi du calcul symbolique, dans la théorie des séries récurrentes*, Nouvelle Correspondance Mathématique, 2 (1876) 201-206, 214.

3. F. Mertens, *Ueber einige asymptotische Gesetze der Zahlentheorie*, Journal für die reine und angewandte Mathematik, 77 (1874) 289-338.

4. G. H. Hardy & E. M. Wright, *An Introduction to the Theory of Numbers*, 3rd ed., Oxford, 1954.

Paul S. Bruckman, Concord, California, concludes his solution with the following remark: By similar methods, we may show that, if P_r denotes the probability that the g.c.d. of two numbers chosen at random from the Fibonacci sequence be equal to r ($r > 1$), then $P_r = 6/(m^2\pi^2)$, where $F_m = r$.

Also solved by Ken Brown, Jean Ezell, Nick Franceschine, Steve Galovitch, Sang-Guen Han, Sidney Heller, W. T. M. Kars (Netherlands), S. C. Locke, O. P. Lossers (Netherlands), L. E. Mattics, Aaron Meyerowitz, Roger B. Nelson, Ernst Trost (Switzerland), L. van Hamme (Belgium), and the proposer.

The Condition $s_n = \Sigma_1^k[a_i n]$

6301 [1980, 494]. *Proposed by Clark Kimberling, University of Evansville.*

Suppose that $\{s_n\}$ is a sequence of positive integers such that

$$0 \leqslant s_{n+m} - s_n - s_m \leqslant K, \quad m, n = 1,2,3,\ldots,$$

for some positive integer K. Let N be a positive integer. Must there exist real numbers $a_1, a_2, \ldots, a_K$ such that

$$s_n = \sum_{k=1}^{K} [a_i n] \quad \text{for } n = 1,2,\ldots,N?$$

(Here $[x]$ denotes the greatest integer less than or equal to x.)

Solution by Andrew D. Pollington, Illinois State University. When $K = 1$ the answer is yes as was shown by Graham, Lin and Lin, "Spectra of Numbers," Math Mag., 51, 174-176. They call such a sequence *nearly linear*. We shall call integer sequences which satisfy the condition

$$0 \leqslant s_{n+m} - s_n - s_m \leqslant K, \quad m, n = 1,2,3,\ldots$$

K-nearly linear sequences. Then problem 6301 becomes "Can every K-nearly linear sequence be decomposed into the sum of K nearly linear sequences?"

When $K > 1$ the answer is no, as is shown by the following counter example. The sequence 1, 2, 4, 6, 7, 8 is 2-nearly linear but cannot be written as the sum of two nearly linear sequences.

This solver then goes on to construct infinitely many counterexamples.

Also solved by Don Coppersmith, J. L. Davison, John Isbell, L. E. Mattics, Martinus Ngantung (Germany), and Howard M. Robbins.

Isbell raises the question: If $s_1, \ldots, s_N$ satisfy the given conditions, can the sequence be continued so that the conditions are still satisfied?

ANSWERS TO "PHOTOS" ON PAGE 331

From top down: B. L. van der Waerden and H. Zassenhaus.

PROBLEMS AND SOLUTIONS

Send all ***proposed*** *problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Random Arcs on a Circle

S 30 [1980, 403]. *Proposed by Solomon W. Golomb, University of Southern California.*

(i) There is a random group of n people in a room. What is the probability that there is a 73-day period (beginning *anywhere* in the year) which contains all their birthdays? (We assume that the birthdays are independent and uniformly distributed modulo 365, we ignore leap-days, and of course the *year* of birth is not counted as part of the birthday.)

(ii) More generally, if n points are placed, independently and at random, on a circle of radius $1/2\pi$ (hence, circumference 1), what is the probability that all these points can be covered by a (movable) arc of length α, $0 \leqslant \alpha \leqslant 1$? Express the answer as a function $P_n(\alpha)$.

Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands.

First solution. It follows directly from W. Feller, *An Introduction to Probability Theory and its Applications II*, 2nd ed., Wiley, Chapter 1, p. 29, Theorem 3, that

$$P_n(\alpha) = 1 - \sum_{\nu=0}^{n} (-1)^{\nu}\binom{n}{\nu}(1-\nu(1-\alpha))_+^{n-1}$$
$$= \sum_{\nu=1}^{n} (-1)^{\nu+1}\binom{n}{\nu}(1-\nu(1-\alpha))_+^{n-1}, \qquad (*)$$

where $(x)_+$ denotes the $\max(0, x)$.

REMARK. For $\alpha \leqslant \frac{1}{2}$ one sees directly that $(*)$ reduces to $P_n(\alpha) = n\alpha^{n-1}$, which can also be proved directly, as our second solution shows.

Second solution.

i) Let p_a denote the probability that $n-1$ persons have their birthday in a given interval of a days.

Clearly

$$p_a = \left(\frac{a}{365}\right)^{n-1}.$$

Given the birthday of the first person of the group we see that the probability that all n persons have their birthday in a 73-day period, equals

$$73p_{73} - 72p_{72} = \frac{73^n - 72^n}{365^{n-1}}. \qquad (**)$$

(Note that a k-day period containing the given day is counted $74 - k$ times by the first term on the LHS and $73 - k$ times by the second term.)

ii) If a year counts D days, then $(**)$ generalizes directly for $a/D \leq \frac{1}{2}$ to

$$a\left(\frac{a}{D}\right)^{n-1} - (a-1)\left(\frac{a-1}{D}\right)^{n-1} = \left(\frac{a}{D}\right)^n D\left(1 - \left(1 - \frac{1}{a}\right)^n\right).$$

So if one keeps $a/D = \alpha$ constant it follows that

$$P_n(\alpha) = \lim_{\substack{D\to\infty \\ a/D=\alpha}} \left(\frac{a}{D}\right)^n D\left(1 - \left(1 - \frac{1}{a}\right)^n\right) = \alpha^n\left(\frac{n}{\alpha}\right) = n\alpha^{n-1}.$$

Notes. M. S. Klamkin commented that both problems, the first in a more general setting, are given and solved elementarily in W. Burnside, *Theory of Probability*, Dover, NY, 1959, pp. 22, 72.

L. Holst suggested his "A note on random arcs on the circle," Technical Report, Department of Statistics, Stanford University, 1980, and H. Solomon's *Geometric Probability*, SIAM CBMS-NSF Series, Philadelphia, 1978, for part of the long and interesting history of (ii).

Also solved by S. W. Dharmadhikari & T. B. Paine, Jordi Dou (Spain), Milton Eisner, David Gootkind, Sidney Heller, Victor Hernandez (Spain), Lars Holst (Sweden), M. S. Klamkin, Roger Pinkham, Arthur Nádas, H. Prodinger (Austria), M. R. Railkar (India), Frederic R. Schwab, Michael Skalsky, Lajos Takács, David M. Wells, and the proposer.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by September 30, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2944. *Proposed by P. Mancevice, Clark University.*

Let $f(x)$ be a positive, continuous, strictly increasing function defined for $x > 0$ such that $f(x) \to \infty$ as $x \to \infty$. Let f^{-1} denote the inverse function. (a) Show that $\Sigma_{n>0} 1/f(n)$ converges if and only if $\Sigma_{n>0} f^{-1}(n)/n^2$ converges. (b) Show that if $\Sigma_{n>0} 1/f(n)$ converges, then

$$\lim_{n\to\infty} n^{-2} \sum_{f(k)\leq n} f(k) = 0.$$

E 2945. *Proposed by H. Kestelman, University College London.*

A and B are $n \times n$ matrices and both have only positive eigenvalues; can AB have only negative eigenvalues? How is the result affected if A, B are both hermitian?

E 2946. *Proposed by Rodica Simon and Frank W. Schmidt, University of Pennsylvania.*

Let $f(n)$ be the product of all positive divisors of n; e.g., $f(3) = 3, f(4) = 8$. Does $f(m) = f(n)$ imply $m = n$?

E 2947. *Proposed by Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Canada.*

Let n and k be integers such that $0 \leqslant k \leqslant n$. Show that

$$\sum_{i=0}^{k}\binom{k}{i}D_{n-i}=\sum_{j=0}^{n-k}(-1)^j\binom{n-k}{j}(n-j)!$$

where $D_m = m!\sum_{r=0}^{m}(-1)^r/r!$ denotes the derangement number (of $1, 2, \ldots, m$).

E 2948. *Proposed by Barry J. Powell, Kirkland, Washington.*

Show that for any fixed pair of coprime positive integers x and y with $x \neq y$, $xy > 1$, there are infinitely many primes p for which the exponent of p in $x^{p-1} - y^{p-1}$ is odd. [The exponent of p in m is the largest integer e such that $p^e \mid m$.]

E 2949. *Proposed by Man-Duen Choi, University of Toronto.*

Let $n_0 < n_1 < \cdots < n_k$ by any given $k+1$ integers. Show by elementary means that the integer $\Pi_{k\geqslant j>i\geqslant 0}(n_j - n_i)$ is exactly divisible by $\Pi_{k\geqslant j>i\geqslant 0}(j-i) = 1!2!\cdots k!$. [An abstract indirect proof of this result has appeared in H. Weyl's book *The Classical Groups*, Chapter VII, Section 5, p. 201, Princeton University Press, 1939.]

SOLUTIONS OF ELEMENTARY PROBLEMS

Superfactorials and Catalan Numbers

E 2799 [1979, 785; 1980, 825]. *Proposed by Marlow Sholander and E. B. Leach, Case Western Reserve University.*

For n a positive integer, let $n!!$ denote the *superfactorial* $\Pi_{i=1}^{n} i!$ and let $0!! = 1$. Set $A_n = (2n-1)!!/[(n-1)!!]^4$. Prove that A_n is an integral multiple of $(2n-1)!$. (A_n is the reciprocal of the determinant of the n by n Hilbert matrix $H_n = (h_{ij})$ with $h_{ij} = (i+j-1)^{-1}$. See Pólya-Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Dover, 1945, vol. 2, Chap. 7, Prob. 3.)

Comment by F. Gerrish, Kingston Polytechnic, England. I confirm that the formula should be:

$$B_n = \frac{1}{(n-1)!}\Pi_{i=1}^{n-1}\binom{n-1+i}{n}\binom{n}{i}.$$

This expression certainly *is* an integer for all integers $n \geqslant 2$; but this has to be proved.

A Criterion that a Geometry be Euclidean

E 2836 [1980, 489]. *Proposed by Joseph E. Valentine, University of Texas at San Antonio.*

Show that an absolute geometry (no parallel postulate) is Euclidean (or Riemannian) if some triangle has the property that a median and the segment joining the midpoints of the other two sides bisect each other.

Solution by Mark D. Meyerson, U.S. Naval Academy, Raymond E. Spaulding, Radford University, and Walter Taylor, University of Colorado (independently). We will assume that the geometry is not Riemannian, so that we may apply all usual betweenness relations (which we will do without further ado); thus we must show that the geometry is Euclidean. In the figure, FD and EB bisect each other at G. By SAS, the four triangles at G are alternately congruent in pairs. Thus the convex quadrilateral $EFBD$ has opposite angles congruent. Moreover $EF = BD = CD$, and likewise $AF = ED$.

The situation can now be summarized by Figures 1 and 2 on page 335.

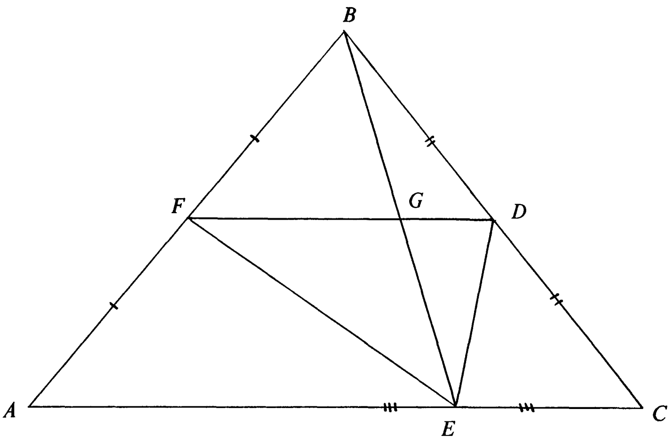


FIG. 1

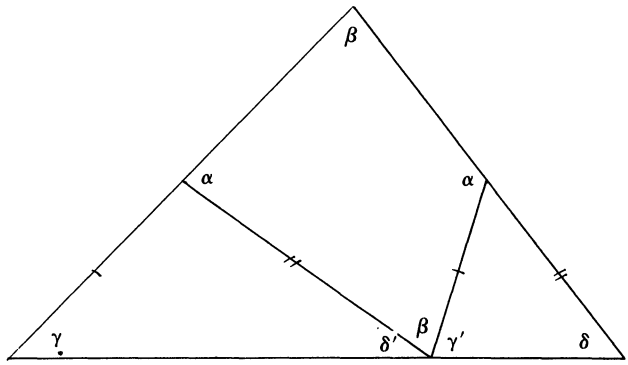


FIG. 2

The two smaller triangles on the base of the large triangle both have apex angle $180 - \alpha$, hence by SAS these two smaller triangles are congruent. Specifically, we have $\gamma = \gamma'$ and $\delta = \delta'$. Thus the angle sum of the large triangle is

$$\gamma + \beta + \delta = \gamma' + \beta + \delta' = 180,$$

and hence our geometry is Euclidean since we have found one triangle of zero defect.

Note. The problem contains a superfluous hypothesis, namely that $AE = EC$. We have written our proof so as to not use this hypothesis.

Also solved by T. Andrei-Dumitou (Rumania), J. E. Carpenter, D. C. Lantz, H. J. Luddwig, and the proposer.

Lantz referred to Martin, *Foundations of Geometry*: Meyerson referred to W. H. Young, *On a form of the parallel axiom*, Quart. J. Math. Oxford Ser. (2), 41 (1910) 353–363.

Minimum Diameter of a Circumscribing Tetrahedron

E 2848 [1980, 671]. *Proposed by Jim Fickett, Texas A & M University.*

Prove that the regular tetrahedron has minimum diameter among all tetrahedra that circumscribe a given sphere. (The diameter is the length of a longest edge.)

Solution by University of South Alabama Problem Group. Let T be a tetrahedron with volume V and surface area A and suppose r is the radius of the sphere inscribed by T. Let one vertex of T be at the origin and let the other three vertices and the center of the sphere be given by the position vectors A, B, C and P respectively. Then $P = \alpha A + \beta B + \delta C$ with $\alpha, \beta, \delta > 0$ and $\alpha + \beta + \delta < 1$.

Suppose that $A \times B$ points into T, then

$$\begin{aligned}
&r|A \times B| = P \cdot (A \times B) = \alpha C \cdot A \times B = 6\alpha V, \text{ and likewise}\\
&r|B \times C| = P \cdot (B \times C) = 6\beta V,\\
&r|C \times A| = P \cdot (C \times A) = 6\delta V,\\
&r|C - A) \times (B - A)| \quad = (P - A) \cdot ((C - A) \times (B - A)) = 6((1 - \alpha) - \beta - \delta)V.
\end{aligned}$$

Solving for α, β, and δ and substituting into the last equation, we find that $r = 3V/A$. The triangle of given perimeter with maximum area is equilateral and (it is possible to show that) the tetrahedron of given surface area with maximum volume is regular. The desired result follows immediately from the last formula.

The solver V. Pambuccian (Rumania) used the formula $V = RS/3$, and referred to N. D. Kazarinoff, *Geometric Inequalities*, 1961, problem 18, pp. 106-108, for a proof of the assertion that the tetrahedron with given surface area and maximum volume is regular. The solver L. Kuipers gave arguments to show that if the longest edge BD exceeds both AD and CD, then there is a circumscribing tetrahedron with smaller diameter.

Also solved by L. E. Mattics.

Covering by Disjoint Closed Intervals

E 2858 [1980, 755]. *Proposed by Oto Strauch, University of Bratislava, Czechoslovakia.*

Let I be the open interval $(0, 1)$, let Y be a countable subset of I, and let a_n $(n \in N)$ be positive real numbers such that $\Sigma_{n=1}^{\infty} a_n \leqslant 1$. Prove that there is a sequence I_n $(n \in N)$ of pairwise disjoint closed subintervals of I such that length $(I_n) = a_n$ for all n, and every point of Y is in the interior of some I_n.

Solution by James Munkres, Massachusetts Institute of Technology.

Two preliminary remarks: (1) The set of limits of subseries of Σa_n is uncountable. Indeed, if Σb_n is a subseries such that $b_{n+1} < b_n/2$ for all n, then any two of *its* subseries have distinct limits. (Consider the smallest n such that b_n appears in one series and not in the other.) (2) We can assume (by rearranging) that $a_{n+1} \leqslant a_n$ for all n.

Let Y be a countable set of reals, indexed with the positive integers.

LEMMA. *Given any interval (a, b) and a series $\Sigma_1^{\infty} b_n \leqslant b - a$, where $0 < b_{n+1} \leqslant b_n$ for all n. Let p be the element of $Y \cap (a, b)$ having smallest index (if any). Then there exists an interval $[c, d]$ of length b_1 contained in (a, b) such that* (i) *$p \in (c, d)$,* (ii) *$c, d \notin Y$, and* (iii) *there is a partition $\{1\} \cup L \cup M$ of the positive integers, with L and M infinite, such that*

$$\Sigma_{n \in L} b_n \leqslant c - a \quad \text{and} \quad \Sigma_{n \in M} b_n \leqslant b - d.$$

Proof. Choose b_0 so that

$$0 \leqslant b_0 < p - a,$$

and

$$p < a + b_0 + \Sigma_1^{\infty} b_n \leqslant b.$$

Let D be the smallest set of reals that is closed under addition and subtraction and contains the point a, each b_n for $n \geqslant 0$, and each point of Y. Then D is countable. Let N be the smallest positive integer such that $p \leqslant a + b_0 + \Sigma_1^N b_n$. Then

$$b_0 + \Sigma_2^N b_n \leqslant b_0 + \Sigma_1^{N-1} b_n < p - a.$$

Choose an infinite set J of even integers greater than N such that

$$\Sigma_{n \in J} b_n < p - a - b_0 - \Sigma_2^N b_n.$$

By (1), there is an infinite set $K \subset J$ such that $\Sigma_{n \in K} b_n$ is not in D. Set

$$c = a + b_0 + \Sigma_2^N b_n + \Sigma_{n \in K} b_n;$$

set $d = c + b_1$. (If $N = 1$, Σ_2^N and Σ_1^{N-1} are defined to be 0.) The conclusion of the lemma holds if we set $L = \{2, \ldots, N] \cup K$, and M to be the complement of $\{1\} \cup L$.

The problem follows readily. Apply the lemma first to the interval $(0, 1)$ and the series Σa_n. Then apply it to the interval $(0, c)$ and the series $\Sigma_{n \in L} a_n$, and to the interval $(d, 1)$ and the series $\Sigma_{n \in M} a_n$. Similarly continue.

Additional comments by Stan Wagon (Smith College). The question arises whether the property of the problem can hold for an uncountable set Y. While we do not know the answer to this, it is consistent that it does not. Borel defined a set of reals to have *strong measure zero* if whenever ε_1, $\varepsilon_2, \ldots$, positive reals, are given, there are disjoint open intervals $J_1, J_2, \ldots$ with length $(J_n) \leqslant \varepsilon_n$, such that the union of these intervals contains Y. It is easy to see that if Y satisfies the conclusion of the problem, then Y has strong measure zero. Moreover, Laver [*On the consistency of Borel's conjecture*, Acta Math., 137 (1976) 151-169] showed that it is consistent that the notion of strong measure zero coincides with countability. We are left with the question of whether the property of the problem coincides with the property of having strong measure zero.

Also solved by J. Henle, E. Johnston, J. G. Mauldon, E. Triesch (West Germany), S. Vaněček (ČSSR), D. M. Wells, W. Zwicker, and the proposer.

Sequence with $a_{mn} \geqslant ma_n$

E 2860 [1980, 823]. *Proposed by J. Martin Borden, University of Illinois.*

Let $\{a_n\}$ $(n = 1, 2, \ldots)$ be a nondecreasing sequence, $0 \leqslant a_{n+1}$. Assume $a_{mn} \geqslant ma_n$ for all m, n, and also $\sup(a_n/n) = c < \infty$. Must a_n/n have a limit?

Solution by Ole Jørsboe, Technical University of Denmark (Lyngby), and Adam Riese, Wright State University, Dayton Ohio. The answer is yes, and the limit is c. If $c = 0$, this is immediate. If $c > 0$, given $\varepsilon > 0$, there must be an N such that $c - \frac{1}{2}\varepsilon < a_N/N \leqslant c$. For every m, it now follows that $c - \frac{1}{2}\varepsilon < a_{mN}/mN \leqslant c$. Whenever $n \geqslant N$ satisfies $mN \leqslant n < (m+1)N$, the relation

$$a_n/n \geqslant a_{mN}/\{(m+1)N\} = [a_{mN}/mN](m/(m+1))$$

holds. But the last expression exceeds $c - \varepsilon$, provided $m/(m+1)$ is sufficiently close to 1. This surely occurs if n is large enough.

Also solved by K. L. Bernstein, R. Breusch, J. A. Cuenca (Spain), J. Fehér (Hungary), L. Harkleroad, V. Hernandez (Spain), G. A. Heuer, E. Johnston, O. P. Lossers (Netherlands), M. R. Modak (India), D. E. Orr, V. Pambuccian (Rumania), E. Posti (Finland), J. P. Robertson, A. Smuckler (Israel), R. A. Struble, E. Triesch (Germany), University of South Alabama Problem Group, and the proposer.

Orthogonal Polynomials of Preassigned Degrees

E 2863 [1981, 66]. *Proposed by D. Shelupsky, The City College, New York, NY.*

Let e_k, $k = 0, 1, 2 \ldots$ be a strictly increasing sequence of real numbers with $e_0 = 0$. Let $p_k(x)$ be the functions derived from the powers x^{e_k} by the Gram-Schmidt process on the interval $[0, 1]$, that is, (a) $p_n(x)$ is a linear combination of x^{e_i}, $0 \leqslant i \leqslant n$, (b) $\int_0^1 p_r(x) p_s(x)\,dx = \delta_{rs}$, and (c) the coefficient c_n of x^{e_n} is positive. Find a formula for c_n.

Solution by Abraham Smuckler, Jerusalem College of Technology. We show that

$$c_n = \sqrt{2e_n + 1} \cdot \prod_{k=0}^{n-1} \left(\frac{1 + e_k + e_n}{e_n - e_k} \right).$$

Our setting is in the Hilbert space $L_2(0,1)$ and we utilize known propositions from there. Let us denote by X_n and Y_n the subspaces spanned respectively by $\{1, x^{e_1}, \dots, x^{e_n}\}$ and $\{p_0, p_1, \dots, p_n\}$. From the Gram-Schmidt construction, $X_n = Y_n$; and, since p_n is orthogonal to p_k for $k < n$, it follows that $p_n \in X_{n-1}^{\perp}$. Now the distance d from the vector x^{e_n} to the closed subspace X_{n-1} in terms of the L_2 norm is: $d = \|x^{e_n} - Px^{e_n}\|$, where Px^{e_n} is the unique vector in X_{n-1} for which the distance d is attained. Px^{e_n} is characterized by the following condition: if $y \in X_{n-1}$, then $y = Px^{e_n}$ if and only if $x^{e_n} - y \in X_{n-1}^{\perp}$. From $p_n = c_0 + c_1 x^{e_1} + \cdots + c_n x^{e_n}$ we have $x^{e_n} - (p_n/c_n) \in X_{n-1}$. Now

$$x^{e_n} - \left(x^{e_n} - \frac{p_n}{c_n} \right) = \frac{p_n}{c_n} \in X_{n-1}^{\perp}$$

and so by the above characterization, $Px^{e_n} = x^{e_n} - p_n/c_n$. It follows that $d = \|p_n/c_n\| = 1/c_n$. But the distance d from x^{e_n} to the subspace spanned by $\{1, x^{e_1}, \dots, x^{e_{n-1}}\}$ is known to be given by:

$$d = \frac{G(x^{e_n}, 1, x^{e_1}, \dots, x^{e_{n-1}})}{G(1, x^{e_1}, \dots, x^{e_{n-1}})}$$

where $G(f_1, f_2, \dots, f_n)$ is the determinant of the Gram matrix $(\int_0^1 f_i \cdot f_j)_{i,j=1}^n$. The known calculation of the above quotient of the two Gram determinants yields:

$$d = \frac{1}{\sqrt{2e_n + 1}} \cdot \prod_{k=0}^{n-1} \left(\frac{e_n - e_k}{1 + e_k + e_n} \right),$$

from which the value of c_n follows.

Also solved by U. Abel (student, Germany), P. S. Bruckman, E. Deutsch, F. Gerrish (England), E. L. Koh (Saudi Arabia), K. M. Levasseur, and the proposer.

Levasseur and Deutsch referred to P. J. Davis' *Interpolation and Approximation*, Dover, 1975, corollary 8.7.6, p. 183.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by September 30, 1982. The solver's full post-office address should be on each sheet.

6386. *Proposed by I. P. Goulden and L. B. Richmond, University of Waterloo.*

Assuming each m-part ordered partition of n to have probability $1/\binom{n-1}{m-1}$, let the expected value of the largest part in such a partition be E_n. Show that, for fixed m, $\lim_{n\to\infty} E_n/n = S_m/m$, where $S_m = \Sigma_1^{m-1} i$ is the mth partial harmonic sum.

6387. *Proposed by Jeff Alden, Vector Research Inc., Ann Arbor, MI.*

Let a, y satisfy $-e^{-1} < a < e^{-1}$, $y = e^{ay}$. Prove that

$$\Sigma_0^\infty n^n a^n / n! = 1/(1 - ay).$$

6388. *Proposed by Nicholas Wheeler and Howard Straubing, Reed College, Portland, OR.*

A regular tetrahedron R sits on a unit triangle T on a plane tiled with triangles congruent to T. A move consists in rotating R about an edge in contact with the plane. After several moves, R sits on T again. Have the vertices of R been permuted in space? What if R is a cube and the tiling is by squares?

6389. *Proposed by Alfonso Castro B., Centro Estudios Avanzados, Mexico, D.F.*

Let Ω denote a generic bounded region in real n-space R^n. Let $H^1(\Omega)$ be the Sobolev space of square integrable functions in Ω having generalized first-order partial derivatives in $L^2(\Omega)$. (See R. Adams, *Sobolev spaces*, Academic Press, 1975.) With integrals over Ω, define $J(u) = [\Sigma_{i=1}^n \int(\partial u/\partial s_i)^2]/\int u^2]$. Find a bounded region Ω such that the infimum (over the nonzero functions u in $H^1(\Omega)$ with $\int u = 0$) is 0.

6390.* *Proposed by P. Venzi and J. Aguadé, ETH, Zürich, Switzerland.*

Let $c_1,\dots,c_n$ be real numbers, with $c_i \neq c_j$ if $i \neq j$. Consider the $n \times n$-matrix (a_{ij}) defined by $a_{ij} = c_j c_i/(c_j - c_i)$ for $i<j$, $a_{ji} = -a_{ij}$ for $i \neq j$ and $\Sigma_{j=1}^n a_{ij} = -c_i$, $i = 1,\dots,n$. Prove that $\det(a_{ij}) = (-1)^n n! c_1 \cdots c_n$.

6391. *Proposed by Louis W. Shapiro, Howard University, Washington, D. C.*

Let a prawn be a new chess piece that moves like a rook in one direction, and like a pawn in the perpendicular direction. A prawn starts in the lower left hand corner of an $n \times n$ board. On each move, it can progress either any number of squares straight up (remaining on the board) or one square to the right.

(a) In how many ways can the prawn get to the mth diagonal (defined by $x + y = m$ on the standard lattice)? This sequence starts 1, 2, 5, 13,....
(b) In how many ways can the prawn get to the opposite corner? This sequence starts 1, 2, 9, 44,....

SOLUTIONS OF ADVANCED PROBLEMS

Relatively Prime Fibonacci Numbers

6300 [1980, 494]. *Proposed by Jack Garfunkel, Flushing, New York.*

Chebycheff showed that, if two numbers are chosen at random from the set of natural numbers, the probability that they will be relatively prime is $6/\pi^2$. Show that if two numbers are chosen at random from the Fibonacci sequence 1, 1, 2, 3, 5,..., the probability P that they will be relatively prime satisfies the inequalities $8/\pi^2 > P > 7/\pi^2$.

Solution by Lajos Takács, Case Western Reserve University. The Fibonacci numbers $u_1 = 1$, $u_2 = 1,\dots,u_n,\dots$ are defined by

$$u_{n+2} = u_{n+1} + u_n \quad \text{for} \quad n \geqslant 1.$$

Denote by $p(x)$ the number of pairs (m,n) satisfying the conditions $1 \leqslant m < n \leqslant x$ and $(u_m, u_n) = 1$. We shall prove that

$$\lim_{x\to\infty} 2p(x)/x^2 = 15/(2\pi^2).$$

If $(m,n) = d$, then by a result of É. Lucas (1876) it follows that $(u_m, u_n) = u_d$. (See also G. H. Hardy & E. M. Wright (1954) p. 148.) Accordingly $(u_m, u_n) = 1$ if and only if either $(m,n) = 1$ or $(m,n) = 2$. Denote by $q(x)$ the number of pairs (m,n) satisfying the conditions $1 \leqslant m < n \leqslant x$ and $(m,n) = 1$. Then the number of pairs (m,n) satisfying the conditions $1 \leqslant m < n \leqslant x$ and $(m,n) = 2$ is $q(x/2)$. Thus $p(x) = q(x) + q(x/2)$.

In 1849 G. Lejeune Dirichlet and in 1874 F. Mertens proved that

$$\lim_{x\to\infty} \frac{2q(x)}{x^2} = \frac{6}{\pi^2}.$$

(See also G. H. Hardy & E. M. Wright (1954) p. 268.) This implies that

$$\lim_{x\to\infty} \frac{2p(x)}{x^2} = \frac{6}{\pi^2}\left(1 + \tfrac{1}{4}\right) = \frac{15}{2\pi^2}.$$

References

1. G. Lejeune Dirichlet, *Über die Bestimmung der mittleren Werthe in der Zahlentheorie*, Abhandlungen der Akademie der Wissenschaften zu Berlin, 1849, pp. 69-83. [G. Lejeune Dirichlet's Werke, Band II, Berlin, 1897, pp. 49-66.]

2. É. Lucas, *Sur l'emploi du calcul symbolique, dans la théorie des séries récurrentes*, Nouvelle Correspondance Mathématique, 2 (1876) 201-206, 214.

3. F. Mertens, *Ueber einige asymptotische Gesetze der Zahlentheorie*, Journal für die reine und angewandte Mathematik, 77 (1874) 289-338.

4. G. H. Hardy & E. M. Wright, *An Introduction to the Theory of Numbers*, 3rd ed., Oxford, 1954.

Paul S. Bruckman, Concord, California, concludes his solution with the following remark: By similar methods, we may show that, if P_r denotes the probability that the g.c.d. of two numbers chosen at random from the Fibonacci sequence be equal to r ($r > 1$), then $P_r = 6/(m^2\pi^2)$, where $F_m = r$.

Also solved by Ken Brown, Jean Ezell, Nick Franceschine, Steve Galovitch, Sang-Guen Han, Sidney Heller, W. T. M. Kars (Netherlands), S. C. Locke, O. P. Lossers (Netherlands), L. E. Mattics, Aaron Meyerowitz, Roger B. Nelson, Ernst Trost (Switzerland), L. van Hamme (Belgium), and the proposer.

The Condition $s_n = \Sigma_1^k[a_i n]$

6301 [1980, 494]. *Proposed by Clark Kimberling, University of Evansville.*

Suppose that $\{s_n\}$ is a sequence of positive integers such that

$$0 \leqslant s_{n+m} - s_n - s_m \leqslant K, \quad m, n = 1,2,3,\ldots,$$

for some positive integer K. Let N be a positive integer. Must there exist real numbers $a_1, a_2, \ldots, a_K$ such that

$$s_n = \sum_{k=1}^{K} [a_i n] \quad \text{for } n = 1,2,\ldots,N?$$

(Here $[x]$ denotes the greatest integer less than or equal to x.)

Solution by Andrew D. Pollington, Illinois State University. When $K = 1$ the answer is yes as was shown by Graham, Lin and Lin, "Spectra of Numbers," Math Mag., 51, 174-176. They call such a sequence *nearly linear*. We shall call integer sequences which satisfy the condition

$$0 \leqslant s_{n+m} - s_n - s_m \leqslant K, \quad m, n = 1,2,3,\ldots$$

K-nearly linear sequences. Then problem 6301 becomes "Can every K-nearly linear sequence be decomposed into the sum of K nearly linear sequences?"

When $K > 1$ the answer is no, as is shown by the following counter example. The sequence 1, 2, 4, 6, 7, 8 is 2-nearly linear but cannot be written as the sum of two nearly linear sequences.

This solver then goes on to construct infinitely many counterexamples.

Also solved by Don Coppersmith, J. L. Davison, John Isbell, L. E. Mattics, Martinus Ngantung (Germany), and Howard M. Robbins.

Isbell raises the question: If $s_1, \ldots, s_N$ satisfy the given conditions, can the sequence be continued so that the conditions are still satisfied?

ANSWERS TO "PHOTOS" ON PAGE 331

From top down: B. L. van der Waerden and H. Zassenhaus.

A Chain of F_σ-sets

6302 [1980, 495]. *Proposed by P. Q. Perlmutter, The Colorado College.*

Let $\{F_\alpha\}_{\alpha\in A}$ be a collection of closed subsets of $\mathbb{R}$. Show that if this collection is a chain, then $F=\cup_{\alpha\in A}F_\alpha$ is an $\mathbb{F}_\sigma$ set. [By a chain, we mean $F_a\subset F_\beta$ or $F_\beta\subset F_a$ for all $\alpha,\beta\in A$.] Can this property be generalized to other classes of Borel sets?

Solution and generalization by F. S. Cater, Portland State University.

THEOREM 1. *Let W be a chain of closed sets in a first countable space. Then $\cup W$ is an F_σ-set. More precisely, $\cup W$ either equals the union of a countable subfamily of W, or $\cup W$ is closed.*

Proof. Suppose that $\cup W$ does not equal the union of a countable subfamily of W. It suffices to prove that $\cup W$ is closed. Assume that $\cup W$ is not closed. Let $x\in\overline{\cup W}$ but $x\notin\cup W$. Let $U_n\,(n=1,2,\ldots)$ be a countable neighborhood system for x. For each n, choose $E_n\in W$ such that $E_n\cap U_n\neq\varnothing$. Then $\cup_n E_n\neq\cup W$, so there is a $G\in W$ such that $G\not\subset\cup_n E_n$. Since W is a chain, $\cup_n E_n\subset G$. So $x\in G$, and $x\in\overline{G}$, contrary to the choice of x. This completes the proof.

REMARK. It is essential that the space be first countable at each point. Consider the order topology on the set of ordinal numbers $\leqslant\Omega$, and let W consist of all closed intervals of the form $[0,a]$ with $a<\Omega$.

Answer to the question posed, by the proposer No. The union of a chain of $\mathbb{F}_\sigma$ sets can be nonmeasurable. Assume *CH*. Let B be some nonmeasurable subset of R. B is uncountable. Let ω_c be the first uncountable ordinal, and well-order B by ω_c, so that $B=\{b_\alpha:\alpha<\omega_c\}$.

Define $B_\alpha=\{b_\delta\in B:\delta<\alpha\}$. Using *CH*, B_α is countable for all $\alpha<\omega_c$. Therefore B_α is an $\mathbb{F}_\sigma$ set.

$$\{B_\alpha\}_{\alpha<\omega_c}\text{ is clearly a chain, and }\cup B_\alpha=B.$$

John C. Morgan II, California State Polytechnic University, Pomona, points out that the solution of the problem for any metric space (less general than Cater's solution) is given by W. Sierpiński, *Sur les ensembles croissantes d'ensembles fermés*, Fund. Math., 36 (1949) 48-50.

Also solved by Teleman Andrei-Dumitru (Romania), Frank B. Miles, Rae Michael Shortt, and the proposer.

Embedding of a Module in a Free Module

6306 [1980, 582]. *Proposed by Joseph Rotman, University of Illinois, Urbana.*

Let R be a commutative ring and M be a finitely generated flat R-module. Can M be imbedded as a submodule of a free R-module?

Solution by A. N. Wiseman, University of Sheffield, England. Not necessarily. Let $R=\Pi_{i=1}^{\infty}R_i$, where each $R_i=\mathbf{Z}_2$, and let I be the ideal of R generated by $e_1,e_2,\ldots$, where e_i is the element with 1 in the ith place and 0 elsewhere. R is a Boolean ring, so the cyclic R-module R/I is flat (since *Boolean* implies *von Neumann regular*). If R/I can be embedded in a free module, then there is a nonzero homomorphism ϕ: $R/I\to R$. Let $\phi(\overline{1})=r_i$), where $\overline{1}=1+I$. For each j, $e_j\overline{1}=\bar{e}_j=0$, so $0=\phi(e_j\overline{1})=e_j(r_i)$ and each $r_i=0$. Hence $\phi(\overline{1})=0$ and $\phi=0$.

There were a number of solutions similar to that given above.

Alberto Facchini, Università di Padova, Italy, and Barbara L. Osofsky, Rutgers University, noted that the answer is "yes," if R is assumed to be an integral domain.

Enzo R. Gentile, Universidad de Buenos Aires, Argentina, makes the following remarks.

1. We needed the following property of von Neumann rings (See: Bourbaki, *Algèbre Commutative*, Chap. I, Modules Plats, Ex. 18 S 2): Every finitely generated submodule of a projective module is a direct summand of it.

2. If R is a Noetherian ring, the contention in the problem is true since for such a ring any f.g. flat module is projective.

3. Commutativity of R plays no role here.

William C. Waterhouse, Pennsylvania State University, gives the following example. R is the ring of locally constant complex functions on the Cantor set and M is the localization MR_I, where I is the ideal of functions vanishing at a chosen point P.

Also solved by Sam Cox, William H. Gustafson, W. K. Nicholson, David E. Rush, Thomas S. Shores, and the proposer.

Injective Modules in a Localization of a Ring

6311 [1980, 675]. *Proposed by Joseph Rotman, University of Illinois, Urbana.*

Let R be a commutative ring and let S be a multiplicatively closed subset of R. If E is an injective R-module, is the localization $S^{-1}E$ an injective $S^{-1}R$-module?

Solution by James Kuzmanovich, Wake Forest University. The localization of an injective module need not be injective. The ring R given below is a self-injective ring with a prime ideal P such that the localization R_P is not self-injective.

Let T be a complete local Noetherian domain with $\dim T \geq 2$, that is, T has a chain of prime ideals $0 \subsetneqq Q \subsetneqq M$ where M is the unique maximal ideal of T. Let E be the T-injective hull of T/M, and let $R = T \times E$ with multiplication defined by $(t, e)(t_1, e_1) = (tt_1, te_1 + t_1e)$. The ring R is self-injective by [1], the T-module E is imbedded as an ideal of R, and $R/E \simeq T$. Let P be the preimage of Q in R and let N be the preimage of M. Consider the localization R_P. If A_n is the annihilator of M^n in E, then by Matlis [2] it follows that $E = \cup_{n=1}^{\infty} A_n$. Take $s \in N - P$. If $x \in E$, then there exists k such that $x \in A_k$ so that $s^k x = 0$. From this it is easily seen that $\ker(R \to R_P) = E$ and that $R_P = T_Q$. The ring R_P thus cannot be self-injective, for it is an integral domain with a nontrivial ideal.

References

1. C. Faith, Self-injective rings, Proc. Amer. Math. Soc., 77 (1979) 157–164.
2. E. Matlis, Injective modules over Noetherian rings, Pacific J. Math., 8 (1958) 511–528.

This problem was also solved by Everett C. Dade; his solution is contained in the note: Localization of injective modules, J. Algebra, 69 (1981) 416–425.

Related problems are discussed in two papers by Zoltan Papp: On stable Noetherian rings, Trans. Amer. Math. Soc., 213 (1975) 107–114; and On the strong injective (projective) dimension of modules, Arch. Math. (Basel) 25 (1974) 354–360.

Gloria Gagola notes that the case that R is Noetherian is given as an exercise on page 163 of I. Kaplansky's *Commutative Rings.*

Complex l_1 Sequences

6319 [1980, 759]. *Proposed by Charles Ryavec, University of California at Santa Barbara.*

Given two complex l_1 sequences $(a_n)_{-\infty}^{\infty}$ and $(b_n)_{-\infty}^{\infty}$ such that $\Sigma a_n = \Sigma b_n$, show that there exists $f \in L^1(R)$ such that $f(n) = a_n$ and $\hat{f}(n) = \int_{-\infty}^{\infty} e^{2\pi i n x} f(x)\, dx = b_n$ for all n. Is the function f unique?

Ignacy Icchak Kotlarski, Oklahoma State University, points out that the problem does not make sense without some further condition on the function f, and provides a solution under the assumption that f is a Fourier original. A shorter existence proof is given by the proposer, under the assumption that f is continuous, which we present below.

$$\text{Put} \quad \Delta(x) = \begin{cases} 1 - x & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geq 1. \end{cases}$$

If $A = \Sigma a_n \neq 0$, define

$$f(x) = \left(\Sigma \frac{b_n}{A} e^{-2\pi i n x}\right)(\Sigma a_n \Delta(x-n));$$

then it is easily verified that f satisfies the conditions. If $A = 0$, let g be the solution obtained by replacing a_0 by $a_0 + 1$ and b_0 by $b_0 + 1$. Then the function $f(x) = g(x) - \Delta(x)$ satisfies the conditions.

Kotlarski points out that if f is a solution, then $f + h$ will also be a solution where, for example, $h(x) = \varepsilon_n \sin 2\pi x$, $x \in [n, n+1)$, for all integers n, with $\Sigma \varepsilon_n = 0$ and $\Sigma |\varepsilon_n| < \infty$.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Theory and Applications of Fourier Analysis. By Charles Sparks Rees, S. M. Shah, and Caslav N. Stanojevic. Marcel Dekker, New York, 1981. 432 pp., $37.50. ISBN 0-8247-6903-1.

ROBERT FEFFERMAN
Department of Mathematics, University of Chicago, Chicago, IL 60637

The central concern of this book is the question of the convergence properties of Fourier series. Roughly speaking, this amounts to asking when it is possible to represent a given 2π-periodic function, $f(\theta)$, $\theta \in R^1$, as a sum

$$f(\theta) = \sum_{n=-\infty}^{+\infty} a_n e^{in\theta}.$$

This question is important in many areas of mathematics, including various parts of analysis, number theory, and geometry. To give one simple example of such an application, consider, in partial differential equations, the Dirichlet problem for the unit disk in R^2. This is the following: Given a function $f(\theta)$, $\theta \in [0, 2\pi)$ solve the equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \text{so that } u(e^{i\theta}) = f(\theta)$$

(i.e., u is to have boundary values $f(\theta)$ on the unit circle). If the given function $f(\theta)$ is $e^{in\theta}$ for some integer n, the solution is immediate: just let $u_n(re^{i\theta}) = r^{|n|} e^{in\theta}$. For a "general" function $f(\theta)$ write $f(\theta)$ as a linear combination of the functions $e^{in\theta}$: $f(\theta) = \Sigma a_n e^{in\theta}$. The solution for this $f(\theta)$ is just the corresponding linear combination of the u_n, i.e., $u = \Sigma_{n=-\infty}^{+\infty} a_n u_n$.

The question of whether or not the Fourier series of a function converges is quite tricky, and the authors have given an account of the most elementary classical results on this topic. For example, if one fixes a single point $\theta_0 \in [0, 2\pi)$, then the continuity of the function is not sufficient to guarantee the convergence of the series at θ_0. In fact there are examples of continuous functions whose Fourier series not only fail to converge at some point θ_0, but whose Fourier partial sums are even unbounded there. However, it turns out that if the function, f, is somewhat smoother at θ_0, say $f'(\theta_0)$ exists, then the Fourier series of f must converge at θ_0 to $f(\theta_0)$. The

reader will find the proof of this result as well as a number of others dealing with related questions presented very clearly in this book.

An interesting feature of this text is that without going into every detail of the construction of Lebesgue measure, the Lebesgue theory is sketched at great length, and the reader without a thorough knowledge of this theory can learn about the integral while immediately seeing why it is so useful.

In addition, the elementary theory of distributions, a few topics in multiple Fourier series, and orthogonal polynomials are discussed, and these are topics not usually found in such an elementary setting. Needless to say, these are very important, and their appearance is quite welcome.

There are comments and exercises after every chapter. They should be taken seriously by the student, because several extremely important elementary concepts, such as the conjugate function, are not treated at all in the body of the book, but only arise briefly in the comments. Several applications of Fourier series (for example, the one mentioned at the beginning of this review) appear in the exercises and these should be of great interest to the student. As a text for a junior or senior undergraduate course, this book is very appealing.

Reduction Methods in Nonlinear Programming. By G. van der Hoek. Mathematisch Centrum, Amsterdam, 1980. pp. iv + 194, ISBN 90-6196-199-8.

D. M. Himmelblau
Department of Chemical Engineering, University of Texas, Austin, TX 78712

Nonlinear programming is the name given to the class of numerical and analytical techniques that solve the problem

$$\text{minimize } f(x) \quad x \in E^n$$

subject to the equality constraints

$$h_i(x) = 0 \quad i = 1, \ldots, m \tag{1}$$

and the inequality constraints

$$g_i(x) \geqslant 0 \quad i = m+1, \ldots, p.$$

In Problem (1) the objective function $f(x)$, the equality constraints $h_i(x)$, and the inequality constraints $g_i(x)$ can be linear or nonlinear. The objective function can be viewed as a performance index, cost function, response surface, or revenue function, while the constraints can be physical limitations, correlations, conservation equations, etc. Excluded from consideration here are those problems which involve integer variables and those problems which include differential or integral equations as constraints. Both these cases require specialized solution techniques.

Methods of solving Problem (1) usually fall into one of two categories: (a) linearization methods, or (b) penalty function (or transformation) methods. Linearization methods are the typical approach to the treatment of nonlinear problems; namely, approximating the nonlinear problem by linearization of the functions (or perhaps using quadratic approximations for some of the functions.) Probably the most obvious form of linearization of Problem (1) is that which leads to a linear programming problem for which many robust computer codes exist.

Penalty function methods encompass a broad class of techniques in which the constraints in Problem (1) are satisfied by adding functionals of them to the objective function $f(x)$ that is being minimized. The concept of a penalty arises from the fact that in minimizing the sum of $f(x)$ and

the functionals, the functionals are devised so that the less the deviation from satisfaction of the constraints, the less the value added to $f(x)$. Exact satisfaction of the constraints would simultaneously lead to no penalty whatsoever and a minimum value for $f(x)$. The reason for the alternate name of transformation methods is that Problem (1) is transformed into an unconstrained problem for which many satisfactory algorithms have been developed. Penalty function methods probably originate with Courant* who suggested studying the relationships between the solutions of a solely equality constrained version of Problem (1) and the solution of unconstrained problems such as

$$\text{minimize } F(x) = f(x) + r_k \sum_{i=1}^{m} h_i^2(x). \tag{1'}$$

Both linearization and penalty function methods use iterative procedures. In linearization methods, the problem is repeatedly approximated by linearization, and an iterative procedure adopted that, one hopes, leads to the solution of the original nonlinear problem by successive stages. Penalty function methods minimize $F(x)$ successively, changing the value of the penalty by modifying the weighting parameter r_k on each cycle of the iteration.

What theory can be brought to bear on Problem (1) is of limited scope: it applies when $F(x)$ is convex and the constraints form a convex set. This means, in practical terms, that the equality constraints must be linear, although the inequality constraints can be nonlinear if the functions are concave. One can find extensive literature devoted to the characteristics of this limited subproblem. Unfortunately, most problems of a scientific and engineering nature that are of interest to practitioners do not meet the convexity requirements, but nevertheless have extrema; one is just not able to prove that such extrema exist.

Improved computer technology has made it possible to seek numerical solutions to Problem (1) via algorithms based on theory, heuristics, or a combination of the two. By far the largest majority of the articles that appear in the literature pertain to algorithms that can be coded for digital computers. Theoretical developments in the last five years appear to be related to the development of algorithms, proof of convergence, determination of rate of convergence (linear, superlinear, quadratic, etc.), or use of special structures. Very few novel concepts which break new ground have appeared. Instead, we find (a) a reinforcement of existing concepts, (b) a filling in of gaps in existing knowledge, and (c) a generalization emerging from originally quite specialized procedures.

Unfortunately, it is not possible to compare algorithms solely on the basis of convergence and rates of convergence because the actual behavior of the algorithm when executed on a computer is definitely problem dependent. In practice, convergence occurs for problems that violate the theoretically imposed convergence conditions so, as expected at the present time, no single nonlinear programming algorithm has proved to be superior to all other nonlinear programming algorithms for every test problem and desired degree of accuracy.

Nevertheless, considerable effort has been exerted in identifying those methods of nonlinear programming that work best on problems of general scope for a substantial bulk of a set of test problems. It is hoped by the investigators that, as in any experimental program, results which have been demonstrated for a certain set of test problems can be expected to be observed for other problems that would be encountered in the future.

As a result of several extensive investigations in the last two or three years, it appears that the successive quadratic approximation method and/or the generalized reduced gradient method are the preferred algorithms to solve Problem (1). Nevertheless, the actual coding of the algorithms exerts a substantial influence upon the effective performance of them so that different versions of these two procedures exist, and they yield different results in computer experimentation. Factors

*R. Courant, Variational methods for the solution of problems of equilibrium and vibrations, Bull. Amer. Math. Soc., 49 (1943) 1–23.

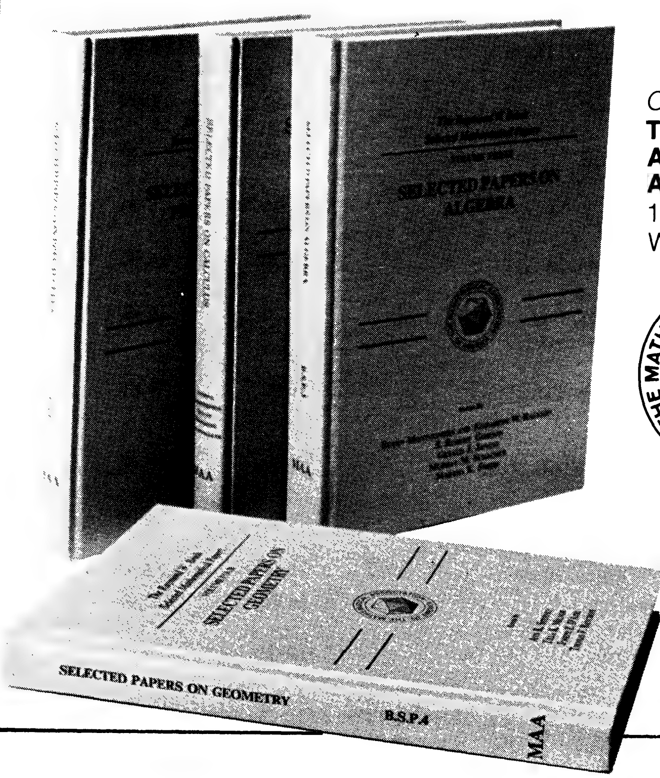
such as the method of handling the matrix manipulations, the avoidance of roundoff and truncation error, the adjustable parameters that can be manipulated by the user, and so on, are all subjective factors that lead to variations in performance and effectiveness of the different algorithms.

Van der Hoek's book focuses on the most general form of the nonlinear programming problem, Problem (1). By reduction method he means transforming Problem (1) into a series of simpler problems in the sense that each subproblem has a lower degree of nonlinearity (not specifically defined) and/or includes fewer constraints. Typical reduction steps might be linearization of the constraints or reduction in the number of degrees of freedom by eliminating variables. Such a definition is useful as a generalization, but too all-encompassing to be of much help in classification.

In the first part of the book, he describes several nonlinear programming procedures in detail, including a number of theorems (many with proofs) that characterize the following types of algorithms: (1) recursive quadratic programming, (2) successive linear programming, (3) penalty function methods. Many of the remarks show considerable insight into the history of the development of each nonlinear programming algorithm and its associated practical problems of implementation.

Specific strategies to execute improved algorithms form the latter part of the book, together with a comparison of computational results for various algorithms in solving 24 test problems. Results for three new algorithms plus four older ones, as executed on the 24 problems, lead to a comparison of robustness and efficiency of the codes. As usual, no single algorithm or code stands out as being *the* best, although the reader can obtain a sense of what might work well for various types of problems from the author's presentation of results and conclusions.

The style and treatment of material in the book are such that it can be read and understood by a nonspecialist without difficulty, but the author also provides a good summary and nice appreciations for a reader experienced in nonlinear programming. The book contains the test problems, references, and an index, but does not include worked-out exercises or homework assignments and study questions, a feature that may limit its use as a classroom text.



SELECTED PAPERS ON GEOMETRY
B.S.P.4
MAA

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N W

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 6 June-July 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

HOW TO BEAT YOUR WYTHOFF GAMES' OPPONENT ON THREE FRONTS

AVIEZRI S. FRAENKEL

Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel 76100

1. Wythoff Games. Let a be a positive integer. Given two piles of tokens, two players move alternately. The moves are of two types: a player may remove any positive number of tokens from a *single* pile, or he may take from both piles, say k (> 0) from one and l (> 0) from the other, provided that $|k - l| < a$. The player first unable to move is the loser, his opponent the winner. Note that passing is not allowed: each player at his turn has to remove at least one token.

We show how to beat our adversary recursively, algebraically and arithmetically. In the course of doing so we shall meet some unexpected and aesthetically pleasing relationships.

The classical Wythoff game [9] is the case $a = 1$, that is, a player taking from both piles has to take the *same* number of tokens from both. See also Coxeter [3]. This special case is reportedly played in China under the name of tsianshidsi. A pleasing presentation of tsianshidsi appears in Yaglom and Yaglom [**10**]. A generalization of the game in a different direction was given by Connell [2]. A generalization including both that of Connell and the one given here is included in [4]. It seems that the more interesting generalization is the one given here, and presenting it alone makes it possible to show what is going on in a more transparent manner.

We start with some notation. Game positions are denoted by (x, y) with $x \leqslant y$, where x denotes the number of tokens in one pile and y the number in the other pile. Positions from which the Previous player can win whatever move his opponent will make, are called *P-positions*, and those from which the Next player can win whatever move his opponent will make are called *N-positions*. Thus $(0,0)$ is a P-position for every a, because the first player is unable to move and so the second player wins; $(0, b)$, $b > 0$, is an N-position for every a; the Next player moves to $(0,0)$ and wins. For $a = 2$, the position $(1,3)$ is a P-position: if Next moves to $(0,3)$, $(0,2)$ or $(0,1)$, then Previous, using a move of the first type, moves to $(0,0)$ and wins. If Next moves to $(1,2)$ or to $(1,1)$, then Previous, using a move of the second type, can again move to $(0,0)$.

The set of all P-positions is denoted by P, and the set of all N-positions by N.

2. A Recursive Characterization of the P-Positions. A list of the first few P-positions for the case $a = 2$ is given in Table 1. The table has an interesting structure. First note that $B_n - A_n = 2n$

TABLE 1. The first few P-positions of Wythoff's game for the case $a = 2$.

n	A_n	B_n
0	0	0
1	1	3
2	2	6
3	4	10
4	5	13
5	7	17
6	8	20
7	9	23
8	11	27
9	12	30
10	14	34

Aviezri Fraenkel has been engaged in communication and computer hardware design, having received his B.Sc. and M.Sc. in electrical engineering. A Ph.D. begun in E.E. ended in 1961 in mathematics under Ernst Straus at U.C.L.A. Since 1962 he has held a position at the Weizmann Institute of Science. He also held visiting positions at a number of American Universities. His research interests lie in the theory of combinatorial games, computational complexity, combinatorial number theory and information retrieval. This and other works of his have been supported materially by his wife Shaula and their six children, one of whom studies mathematics and another considers an engineering career.

(an in general). It is probably a bit harder to notice that $A_n = \operatorname{mex}\{A_i, B_i : i < n\}$, where, for any set S, if $\bar{S}$ denotes the complement of S with respect to the nonnegative integers, then $\operatorname{mex} S = \min \bar{S} =$ least nonnegative integer not in S. (mex stands for *m*inimum *ex*cluded value. The term has been coined by John H. Conway, I believe.) Thus $\operatorname{mex} \varnothing = 0$. If we define the pairs (A_n, B_n) in the indicated manner for all n, then $(A_{11}, B_{11}) = (15, 37)$, since 15 is the smallest nonnegative integer not yet in the table.

We now prove formally that the pairs (A_n, B_n) as defined above do indeed constitute the set of P-positions of the game.

THEOREM 1. *$P = \cup_{i=0}^{\infty}\{(A_i, B_i)\}$, where $A_n = \operatorname{mex}\{A_i, B_i : 0 \leqslant i < n\}$ and $B_n = A_n + an$ $(n \geqslant 0)$.*

Proof. From the definition of A_n and B_n as given in the theorem it follows that if $A = \cup_{n=1}^{\infty} A_n$ and $B = \cup_{n=1}^{\infty} B_n$, then A and B are *complementary* sets of numbers, that is, $A \cup B = Z^+$ (the set of positive integers), and $A \cap B = \varnothing$. The last equality is true since if $A_n = B_m$, then $n > m$ implies that A_n is the mex of a set containing $B_m = A_n$, a contradiction; and $n \leqslant m$ is impossible since $B_m = A_m + am \geqslant A_n + an > A_n$.

In order to prove the theorem it evidently suffices to show two things: I. A player moving from some (A_n, B_n) lands in a position not of the form (A_i, B_i). II. Given any position $(x, y) \neq (A_i, B_i)$, there is a move to some (A_n, B_n). (It is useful to note that these two conditions are also necessary: the definition of P and N implies that *all* positions reachable in one move from a P-position are N-positions; whereas at least one P-position is reachable in one move from an N-position.)

I. A move of the first type from (A_n, B_n) clearly leads to a position not of the form (A_i, B_i). Suppose that a move of the second type from (A_n, B_n) produces a position (A_i, B_i). Then $i < n$. A move of the second type satisfies $|(B_n - B_i) - (A_n - A_i)| < a$, that is, $|(n - i)a| < a$, which implies $i = n$, a contradiction.

II. Let (x, y) with $x \leqslant y$ be a position not of the form (A_i, B_i) $(i \geqslant 0)$. Since A and B are complementary, every positive integer appears exactly once in exactly one of A and B. Therefore we have either $x = B_n$ or else $x = A_n$ for some $n \geqslant 0$.

Case (i): $x = B_n$. Then move $y \to A_n$.

Case (ii): $x = A_n$. If $y > B_n$, then move $y \to B_n$. If $A_n \leqslant y < B_n$, let $d = y - x$, $m = \lfloor d/a \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer $\leqslant x$. Then move $(x, y) \to (A_m, B_m)$. This is a legal move, since:

(a) $d = y - A_n < B_n - A_n = an$, hence $m = \lfloor d/a \rfloor \leqslant d/a < n$,

(b) $y = A_n + d > A_m + d \geqslant A_m + am = B_m$,

(c) $|(y - B_m) - (x - A_m)| = |(y - x) - (B_m - A_m)| = |d - am| < a$. ∎

In order to play a game such as a Wythoff game as best as possible, it suffices to compute two things: (A) the nature of the present position u (P or N); (B) a next move if u is in N. Reason: let u be an artibrary game position. If (A) shows that $u \in N$, then we know that there exists some move to a position in P. Moreover, we can use (B) to find one. If, on the other hand, (A) shows that $u \in P$, we cannot do much better than an arbitrary move while exuding an air of confidence and hoping for the best, since any position reachable in one move from a P-position is necessarily an N-position, from which our opponent can win if he knows to compute (A) and (B). Now the *statement* of Theorem 1 shows how to compute (A), since the statement constitutes a characterization of the P-positions, whereas the *proof* of Theorem 1 indicates explicitly how to compute (B).

The computation of (A) and (B) (or of (B) alone when the computation of (A) is already known) will be called a *strategy* in the sequel. Summarizing our present knowledge, we can thus say that Theorem 1 and its proof jointly constitute a recursive strategy for Wythoff games in which each P-position can be computed from the previous ones.

We close this section by briefly considering the complexity of the indicated strategy. Given a

position (x, y) with $0 \leqslant x \leqslant y$, we need, for computing the next move, to construct the table recursively only up to the smallest n such that either $A_n = x$ or $B_n = x$. Since $A_n \leqslant 2n$ for every a (follows from $B_n - B_{n-1} \geqslant 2$), this computation requires only $O(x)$ comparisons of table entries with x, and $O(x)$ words of memory space. We remark that once the table has been computed and stored, it takes only $O(\log x)$ steps to locate A_n such that $x = A_n$ (or B_n such that $x = B_n$), by performing a binary search in the A_n (or B_n) sequence. Since computing (B) by the method indicated in the proof of Theorem 1 requires at most $O(\log x)$ steps, the total number of computation steps is only $O(x)$. In the next section we give a closed form for the nth P-position, which enables us to beat our adversary using an explicit rather than only an implicit recursive construction, which is at the same time computationally more efficient!

3. An Algebraic Characterization of the *P*-Positions. Let

$$\alpha = \frac{2 - a + \sqrt{a^2 + 4}}{2}, \quad \beta = \alpha + a. \tag{1}$$

α is the positive root of the quadratic equation $\xi^{-1} + (\xi + a)^{-1} = 1$. Thus α and β are irrational for every positive integer a, and satisfy $\alpha^{-1} + \beta^{-1} = 1$.

The following "folk-theorem" dates back at least to Beatty [**1**]. It has many proofs and has often been rediscovered. The proof given below seems to be one of the most elegant ones. I have heard that it is due to Ostrowski.

LEMMA 1. *Let α and β be positive irrationals satisfying $\alpha^{-1} + \beta^{-1} = 1$. Let $A'_n = \lfloor n\alpha \rfloor$, $B'_n = \lfloor n\beta \rfloor$ $A' = \cup_{n=1}^{\infty}\{A'_n\}$ and $B' = \cup_{n=1}^{\infty}\{B'_n\}$. Then A' and B' are complementary.*

Proof. It suffices to show that exactly one member of the sequence $\zeta = \{\alpha, \beta, 2\alpha, 2\beta, 3\alpha, 3\beta, \ldots, n\alpha, n\beta, \ldots\}$ is in the interval $[h, h+1)$ for every positive integer h. Hence it suffices to show that if $M > 1$ is any integer, then there are precisely $M - 1$ members of ζ less than M. The number of $n\alpha < M$ is $\lfloor M/\alpha \rfloor$ and the number of $n\beta < M$ is $\lfloor M/\beta \rfloor$. Thus the number of members of ζ less than M is $N = \lfloor M/\alpha \rfloor + \lfloor M/\beta \rfloor$. Now

$$\frac{M}{\alpha} - 1 < \left\lfloor \frac{M}{\alpha} \right\rfloor < \frac{M}{\alpha}, \frac{M}{\beta} - 1 < \left\lfloor \frac{M}{\beta} \right\rfloor < \frac{M}{\beta}.$$

Adding, $M - 2 < N < M$. Since N is an integer, we conclude $N = M - 1$. ∎

Note that $A'_0 = 0 = A_0$, $B'_0 = 0 = B_0$ and $B'_n = A'_n + an$. Moreover, $\operatorname{mex}\{A'_i, B'_i : 0 \leqslant i < n\} = A'_n$ $(n \geqslant 0)$, since A'_n and B'_n are increasing sequences and A' and B' are complementary: if the mex were not A'_n, then A'_n would never be obtained! This shows that $A'_n = A_n$ and $B'_n = B_n$ $(n \geqslant 0)$. We have proved:

THEOREM 2. *If α and β are given by* (1), *then $P = \cup_{n=0}^{\infty}\{(\lfloor n\alpha \rfloor, \lfloor n\beta \rfloor)\}$.*

A strategy based on this observation can be realized as follows. Since α is irrational and $1 < \alpha < 2$,

$$x = \lfloor n\alpha \rfloor \Leftrightarrow x < n\alpha < x + 1 \Leftrightarrow \frac{x}{\alpha} < n < \frac{x+1}{\alpha} \Leftrightarrow \left\lfloor \frac{x+1}{\alpha} \right\rfloor = \left\lfloor \frac{x}{\alpha} \right\rfloor + 1,$$

where (x, y) with $x \leqslant y$ is a given game position. Therefore either $x = \lfloor n\alpha \rfloor = A_n$ where $n = \lfloor (x+1)/\alpha \rfloor$, or else, by complementarity, $x = \lfloor n\beta \rfloor = B_n$, where $n = \lfloor (x+1)/\beta \rfloor$. We have thus reduced the situation to that considered in cases (ii) and (i) in the proof of Theorem 1, and hence the strategy presented in that proof can be followed. For example, if $x = \lfloor n\alpha \rfloor = A_n$ and $y < \lfloor n\alpha \rfloor + na = \lfloor n\beta \rfloor$, then letting $m = \lfloor (y - x)/a \rfloor$, we move to $(\lfloor m\alpha \rfloor, \lfloor m\beta \rfloor) \in P$. For implementing this strategy, α has to be computed to a precision of $O(\log x)$ digits, and its storage requires $O(\log x)$ words, which is only the same order of magnitude needed for storing x itself (in binary or decimal, say).

In order to give still another unexpected way for beating our opponent, we resort to the theory of continued fractions.

4. Continued Fractions and Systems of Numeration. Let α be an irrational number satisfying $1 < \alpha < 2$. Denote its *simple continued fraction* expansion by

$$\alpha = 1 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}} = [1, a_1, a_2, a_3, \ldots],$$

where the a_i are positive integers. Its *convergents* $p_n/q_n = [1, a_1, \ldots, a_n]$ satisfy the recursion

$$p_{-1} = 1,\ p_0 = 1,\ p_n = a_n p_{n-1} + p_{n-2} \qquad (n \geqslant 1)$$
$$q_{-1} = 0,\ q_0 = 1,\ q_n = a_n q_{n-1} + q_{n-2} \qquad (n \geqslant 1).$$

We do not need much more on continued fractions. The reader who wishes to read up on the theory of continued fractions may like to consult, for example, Hardy and Wright [**5**, Ch. 10], Olds [**7**] or Perron [**8**]. What we do need is the fact that every irrational number has a unique expansion into an infinite continued fraction and that conversely, every infinite continued fraction represents a unique irrational number. Moreover, we will use the fact that $\alpha = [1, \dot{a}]$ where α is given by (1) and $\dot{a}$ denotes the infinite repetition of a, and a property stated just prior to Lemma 3 below.

In the next theorem we give two systems of numeration, one based on the numerators p_i and one on the denominators q_i of the convergents of α. The two systems are called *p-system* and *q-system* in the sequel.

THEOREM 3. *Every positive integer can be written uniquely in the form*

$$N = \sum_{i=0}^{m} s_i p_i,\ 0 \leqslant s_i \leqslant a_{i+1},\ s_{i+1} = a_{i+2} \Rightarrow s_i = 0 \quad (i \geqslant 0), \tag{2}$$

and also in the form

$$N = \sum_{i=0}^{n} t_i q_i,\ 0 \leqslant t_0 < a_1,\ 0 \leqslant t_i \leqslant a_{i+1},\ t_i = a_{i+1} \Rightarrow t_{i-1} = 0 \quad (i \geqslant 1). \tag{3}$$

Note. Putting $a_i = 1$ $(i \geqslant 1)$, (2) becomes the *Fibonacci counting system*, in which all the digits s_i are 0 or 1. This is the usual binary numeration system, except that there are never two consecutive ones. This system is discussed, e.g., in Knuth [**6**, Sect. 1.2.8, Ex. 34] and in Yaglom and Yaglom [**10**].

Table 2 displays the representation of the first few positive integers in the p and q-systems for the case $a_i = 2$ $(i \geqslant 1)$.

Proof. We shall prove the result for the p-system. The proof for the q-system is very similar. Given a positive integer N, let m be the largest integer such that $p_m \leqslant N$. Write

$$\begin{aligned}
N &= s_m p_m + r_m, & 0 &\leqslant r_m < p_m\\
r_m &= s_{m-1} p_{m-1} + r_{m-1}, & 0 &\leqslant r_{m-1} < p_{m-1}\\
&\ \vdots\\
r_{i+1} &= s_i p_i + r_i, & 0 &\leqslant r_i < p_i\\
&\ \vdots\\
r_2 &= s_1 p_1 + r_1, & 0 &\leqslant r_1 < p_1\\
r_1 &= s_0 p_0.
\end{aligned}$$

TABLE 2. The representation of the first few positive integers in the p and q-systems for the case $a_i = 2$ $(i \geqslant 1)$

q_3	q_2	q_1	q_0	p_3	p_2	p_1	p_0	
12	5	2	1	17	7	3	1	n
			1				1	1
		1	0				2	2
		1	1			1	0	3
		2	0			1	1	4
	1	0	0			1	2	5
	1	0	1			2	0	6
	1	1	0		1	0	0	7
	1	1	1		1	0	1	8
	1	2	0		1	0	2	9
	2	0	0		1	1	0	10
	2	0	1		1	1	1	11
1	0	0	0		1	1	2	12
1	0	0	1		1	2	0	13
1	0	1	0		2	0	0	14
1	0	1	1		2	0	1	15
1	0	2	0		2	0	2	16
1	1	0	0	1	0	0	0	17

Thus

$$N = \sum_{i=0}^{m} s_i p_i, \tag{4}$$

that is, N is representable in the p-system. (The careful reader will note that up to this point we have not used properties of continued fractions. Thus if $1 = p_0 < p_1 < p_2 < \cdots$ is any sequence of positive integers, then the representation (4) holds. Letting, e.g., $p_i = b^i$ $(b > 1$ fixed) leads to the usual representation of N to the base b.) The digits s_i of the representation (4) satisfy

$$s_i = \frac{r_{i+1} - r_i}{p_i} < \frac{p_{i+1}}{p_i} = \frac{a_{i+1}p_i + p_{i-1}}{p_i} = a_{i+1} + \frac{p_{i-1}}{p_i} \leqslant a_{i+1} + 1,$$

and so $0 \leqslant s_i \leqslant a_{i+1}$ $(i \geqslant 0)$. (However, since $q_{-1} = 0$, we get $t_0 < a_1$ for the q-system.) Suppose that $s_i = a_{i+1}$ and $s_{i-1} \geqslant 1$. Then

$$r_i \geqslant p_{i-1} \text{ and so } r_{i+1} \geqslant a_{i+1}p_i + p_{i-1} = p_{i+1},$$

a contradiction. Hence $s_i = a_{i+1} \Rightarrow s_{i-1} = 0$ $(i \geqslant 1)$.

For proving uniqueness we need the following auxiliary result.

LEMMA 2. *Let*

$$H_{i+1} = a_{i+1}p_i + a_{i-1}p_{i-2} + \cdots + a_{k+1}p_k,$$

where $k = 0$ if i is even, $k = 1$ if i is odd. Then $H_{i+1} = p_{i+1} - 1$.

In other words, the expression H_{i+1} is the equivalent in the p-system of 99....9 in the decimal system.

Proof. We add $p_{i-1} - p_{i-1}$ to the first term, $p_{i-3} - p_{i-3}$ to the second, etc. Using $p_n = a_n p_{n-1} + p_{n-2}$ $(n \geqslant 1)$, we then get

$$H_{i+1} = (p_{i+1} - p_{i-1}) + (p_{i-1} - p_{i-3}) + \cdots \begin{cases} +(p_1 - 1) & (i \text{ even}) \\ +(p_2 - p_0) & (i \text{ odd}). \end{cases}$$

Since $p_0 = 1$, we get in both cases $H_{i+1} = P_{i+1} - 1$. ■

We now resume the proof of Theorem 3. Suppose that N has two different representations:

$$N=\sum_{i=0}^{m} s_i p_i=\sum_{i=0}^{m} u_i p_i,$$

where the digits s_i and u_i satisfy the conditions imposed in (2). Let j be the largest integer such that $s_j \neq u_j$, say $s_j > u_j$. Then

$$\begin{aligned} p_j \leqslant (s_j-u_j)p_j &= \sum_{i=0}^{j-1}(u_i-s_i)p_i \leqslant \sum_{i=0}^{j-1} u_i p_i \\ &\leqslant \sum_{i=0}^{\lfloor (j-1)/2 \rfloor} a_{j-2i}p_{j-2i-1}=p_j-1, \end{aligned}$$

a contradiction. (The last equality in this chain is Lemma 2, and the inequality just preceding it follows from the identity

$$a_{i+1}p_i > (a_{i+1}-1)p_i + a_i p_{i-1} \quad (i \geqslant 1).) \qquad \blacksquare$$

We close this section with two definitions which will be useful in the next and final section.

(i) Relative to a simple continued fraction $\alpha=[1, a_1, a_2, \ldots]$, we define a *representation* R to be an $(m+1)$-tuple

$$R=(d_m, d_{m-1}, \ldots d_1, d_0),$$

where

$$0 \leqslant d_i \leqslant a_{i+1} \quad \text{and} \quad d_{i+1}=a_{i+2} \Rightarrow d_i=0 \quad (i \geqslant 0).$$

If it is known that $d_{i-1}=d_{i-2}=\cdots=d_0=0$, we also write $R=(d_m,\ldots,d_i)$ instead of $(d_m,\ldots,d_i,0,\ldots,0)$. The *p-interpretation* I_p of a representation $R=(d_m,\ldots,d_0)$ is the number $I_p=\sum_{i=0}^{m} d_i p_i$. The *q-interpretation* of R is the number $I_q=\sum_{i=0}^{m} d_i q_i$, provided that $d_0 < a_1$; otherwise there is no q-interpretation of R. Given any positive integer k, we say that its *p-representation* $R_p(k)$ (or *q-representation* $R_q(k)$) is $(d_m,\ldots,d_0)$ if

$$k=\sum_{i=0}^{m} d_i p_i \quad \left(\text{or} \quad k=\sum_{i=0}^{m} d_i q_i,\, d_0 < a_1\right).$$

We shall later be interested in p-interpretations of q-representations! Thus for $a_i=2$ $(i \geqslant 1)$, the decimal number 12 has q-representation 1000 (see Table 2), whose p-interpretation is 17. In other words, $I_p(R_q(12))=I_p(1000)=17$.

(ii) If $R=(d_m,\ldots,d_0)$ is any representation (which might be $R_p(k)$ or $R_q(k)$ for some positive integer k), then the representation $R'=(d_m,\ldots,d_0,0)$ is called a *left shift* of R. In other words, R' is obtained from R by shifting each digit d_i of R left by one place and inserting a zero at the right. If $R=(d_m,\ldots,d_1,d_0)$ is a representation with $d_0=0$, then the representation $R''=(d_m,\ldots,d_1)$ is called a *right shift* of R.

5. An Arithmetic Characterization of the *P*-positions. We use the new numeration system introduced in the last section to give yet another, quite different, characterization of the P-positions.

Comparing Tables 1 and 2 we notice three interesting patterns. In Theorems 4 and 5 below we show that they hold indeed, in general, in the form of the following three properties.

Property 1. The set of numbers $A_n=\operatorname{mex}\{A_i, B_i : 0 \leqslant i < n\}$ $(n \geqslant 1)$ is identical to the set of numbers with p-representations ending in an even number of zeros; and the set of numbers $B_n=A_n+an$ $(n \geqslant 1)$ is identical to the set of numbers with p-representations ending in an odd number of zeros. Thus in Table 2 $R_p(1)$ ends in an even number (zero) of zeros, and so does $R_p(7)$ (ending in two zeros). Both 1 and 7 are in the A_n-column of Table 1.

Property 2. For every $n \geqslant 1$, the p-representation of B_n is a left shift of A_n: $R_p(B_n)=R'_p(A_n)$.

(Thus $(1,3)$ of Table 1 has p-representation $(1,10)$, and $(5,13)$ has p-representation $(12,120)$.)

Property 3. Let n be any positive integer. If $R_q(n)$ ends in an even number of zeros, then $I_p(R_q(n)) = A_n$. (Thus for $a_i = 2$ $(i \geq 1)$, $I_p(R_q(5)) = I_p(100) = 7 = A_5$.) If $R_q(n)$ ends in an odd number of zeros, then $I_p(R_q(n)) = A_n + 1$. (Thus for the case above, $I_p(R_q(4)) = I_p(20) = 6 = A_4 + 1$.)

For proving these properties we need an auxiliary result. Let $\alpha = [1, a_1, a_2, \ldots]$ be irrational with convergents $\{p_i/q_i\}$. Let $D_i = \alpha q_i - p_i$ $(i \geq 1)$. From the theory of continued fractions it is known that

$$-1 = D_{-1} < D_1 < D_3 < \cdots < 0 < \cdots < D_4 < D_2 < D_0 = \alpha - 1.$$

LEMMA 3. $D_j + \sum_{i=1}^{m} a_{j+2i} D_{j+2i-1} = D_{j+2m} \quad (j \geq -1)$.

Proof. We have

$$D_j + a_{j+2} D_{j+1} = \alpha q_j - p_j + a_{j+2}(\alpha q_{j+1} - p_{j+1}) = D_{j+2}$$

Thus

$$\begin{gathered} D_j + a_{j+2} D_{j+1} = D_{j+2}, \\ D_{j+2} + a_{j+4} D_{j+3} = D_{j+4}, \\ \vdots \\ D_{j+2m-2} + a_{j+2m} D_{j+2m-1} = D_{j+2m}. \end{gathered}$$

Adding proves the assertion. ■

The proof of Property 3 follows from the next theorem.

THEOREM 4. *Let $\alpha = [1, a_1, a_2, \ldots]$ be irrational with convergents $\{p_i/q_i\}$. Let n be a positive integer. If $R_q(n) = (d_m, \ldots, d_{2k})$ $(d_{2k} \neq 0, k \geq 0)$, then $I_p(R_q(n)) = \lfloor n\alpha \rfloor$. (That is,*

$$n = \sum_{i=2k}^{m} d_i q_i \Rightarrow \lfloor n\alpha \rfloor = \sum_{i=2k}^{m} d_i p_i.)$$

If $R_q(n) = (d_m, \ldots, d_{2k+1})$ $(d_{2k+1} \neq 0, k \geq 0)$, then $I_p(R_q(n)) = \lfloor n\alpha \rfloor + 1$. (That is,

$$\begin{aligned} n = \sum_{i=2k+1}^{m} d_i q_i \Rightarrow \lfloor n\alpha \rfloor &= -1 + \sum_{i=2k+1}^{m} d_i p_i \\ &= \sum_{i=0}^{k} a_{2i+1} p_{2i} + (d_{2k+1} - 1) p_{2k+1} + \sum_{i=2k+2}^{m} d_i p_i, \end{aligned}$$

where the last equality follows from Lemma 2.)

Proof. For the first case it suffices to show that

$$0 < n\alpha - \sum_{i=2k}^{m} d_i p_i < 1,$$

that is,

$$0 < \sum_{i=2k}^{m} d_i D_i < 1.$$

By Lemma 3,

$$\sum_{i=2k}^{m} d_i D_i \geq D_{2k} + \sum_{i=1}^{m} a_{2k+2i} D_{2k+2i-1} = D_{2k+2m} > 0,$$

$$\sum_{i=2k}^{m} d_i D_i \leqslant \sum_{i=1}^{m} a_{2k+2i-1} D_{2k+2i-2} = D_{2k+2m-1} - D_{2k-1}$$
$$\leqslant D_{2k+2m-1} + 1 < 1.$$

For the second case it suffices to show that

$$-1 < \sum_{i=2k+1}^{m} d_i D_i < 0.$$

Again by Lemma 3,

$$\sum_{i=2k+1}^{m} d_i D_i \geqslant \sum_{i=1}^{m} a_{2k+2i} D_{2k+2i-1} = D_{2k+2m} - D_{2k}$$
$$\geqslant -D_{2k} \geqslant -D_0 = 1 - \alpha > -1,$$
$$\sum_{i=2k+1}^{m} d_i D_i \leqslant D_{2k+1} + \sum_{i=1}^{m} a_{2k+2i+1} D_{2k+2i} = D_{2k+2m+1} < 0. \qquad \blacksquare$$

We now prove Property 2.

THEOREM 5. *Let $\alpha = [1, \dot{a}]$, $\beta = \alpha + a$, where a is any positive integer. Then for every positive integer n, $R_p(\lfloor n\beta \rfloor) = R_p'(\lfloor n\alpha \rfloor)$.*

Proof. We have $\lfloor \alpha \rfloor = 1 = p_0$, $\lfloor \beta \rfloor = 1 + a = p_1$; hence the claim holds for $n = 1$. Suppose it holds for all $k < n$. Now $R_p(\lfloor n\alpha \rfloor)$ ends in an even number of zeros by Theorem 4. Let R' be the left shift of $R_p(\lfloor n\alpha \rfloor)$. By the induction hypothesis, $I_p(R') \neq \lfloor k\beta \rfloor$, $k < n$. In fact, $I_p(R')$ is the smallest number with representation R' ending in an odd number of zeros not yet obtained. If $I_p(R') \neq \lfloor n\beta \rfloor$, then $I_p(R')$ can never be obtained for $k > n$, since the sequence $\lfloor k\beta \rfloor$ is increasing, in contradiction to Lemma 1. $\blacksquare$

Theorem 4 asserts that $R_p(\lfloor n\alpha \rfloor)$ ends in an even number of zeros for all n. Theorem 5 implies, in particular, that $R_p(\lfloor n\beta \rfloor)$ ends in an odd number of zeros for all n. Since the sequences $\lfloor n\alpha \rfloor$ and $\lfloor n\beta \rfloor$ are complementary, every positive integer k such that $R_p(k)$ ends in an even (odd) number of zeros has the form $\lfloor n\alpha \rfloor$ ($\lfloor n\beta \rfloor$). This proves Property 1.

Now suppose we are given a position (x, y) with $0 < x \leqslant y$. To obtain a strategy based on Theorems 4 and 5, we compute $R_p(x)$. If it ends in an odd number of zeros, then $x = B_k$ for some k, and a winning move is $(x, y) \to (I_p(R_p''(x)), x) \in P$. If $R_p(x)$ ends in an even number of zeros, then $x = A_k$ for some k. If $y > I_p(R_p'(x))$, then the move $(x, y) \to (x, I_p(R_p'(x))) \in P$ is a winning move. If $y = I_p(R_p'(x))$, then $(x, y) \in P$, so we cannot win when starting from the given position (x, y). Finally, if $x \leqslant y < I_p(R_p'(x))$, then let $m = \lfloor (y - x)/a \rfloor$. If $R_q(m)$ ends in an even number of zeros, then $I_p(R_q(m)) = A_m$ by Theorem 4. If $R_q(m)$ ends in an odd number of zeros, then $I_p(R_q(m)) = A_m + 1$. In either case, a winning move is $(x, y) \to (A_m, A_m + ma) \in P$.

In order to estimate the complexity of this algorithm note that for $\alpha = \alpha_1 = [1, \dot{a}]$, the solution of the recursion $p_{-1} = 1$, $p_0 = 1$, $p_n = ap_{n-1} + p_{n-2}$ $(n \geqslant 1)$ is

$$p_n = \frac{1}{\alpha_1 - \alpha_2}\left(\alpha_1(\alpha_1 + a - 1)^{n+1} - \alpha_2(\alpha_2 + a - 1)^{n+1}\right)$$
$$= \frac{1}{\sqrt{a^2+4}}\left(\alpha_1\left(\frac{a + \sqrt{a^2+4}}{2}\right)^{n+1} - \alpha_2\left(\frac{a - \sqrt{a^2+4}}{2}\right)^{n+1}\right),$$

where

$$\alpha_1 = \frac{2 - a + \sqrt{a^2+4}}{2}, \qquad \alpha_2 = \frac{2 - a - \sqrt{a^2+4}}{2}$$

are the two roots of the quadratic equation $\xi^{-1} + (\xi + a)^{-1} = 1$. Since $-1 < \alpha_2 + a - 1 < 0$, we have $p_n \sim cg^{n+1}$, where $c = \alpha_1/\sqrt{a^2+4}$ and $g = (a + \sqrt{a^2+4})/2$.

Let n be the largest integer such that $x \geqslant p_n$. Since $p_n \sim cg^{n+1}$ we see that $n = O(\log x)$ steps suffice to compute $R_p(x)$ and $O(\log x)$ words of memory suffice to store it, since for computing $R_p(x)$ we need only the first $O(\log x)$ values of p_i. For the case $A_n = x \leqslant y < B_n$, since $m = \lfloor (y-x)/a \rfloor < n$, the computation of $R_q(m)$ also requires at most $O(\log x)$ steps and that much memory space. Since $n \sim \log_g(x/c)$ and g increases with a, it is seen that for large a this strategy implementation is more efficient than even the algebraic one of the previous section.

Acknowledgement. The two referees unanimously and independently purged a second part of this paper which constituted the author's sole motivation for writing the article in the first place! Nevertheless (and since both referees unanimously and independently recommended publishing the second part elsewhere), the author wishes to thank them and the editor for their very useful, detailed and constructive comments which helped to improve the article in various ways.

References

1. S. Beatty, Problem 3173, this MONTHLY, 33 (1926) 159; 34 (1927) 159.
2. I. G. Connell, A generalization of Wythoff's game, Canad. Math. Bull., 2 (1959) 181–190.
3. H. S. M. Coxeter, The golden section, Phyllotaxis and Wythoff's game, Scripta Math., 19 (1953) 135–143.
4. A. S. Fraenkel and I. Borosh, A generalization of Wythoff's game, J. Comb. Theory, 15 (1973) 175–191.
5. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 4th ed., 1960.
6. D. E. Knuth, The Art of Computer Programming, vol. 1: Fundamental Algorithms, Addison-Wesley, Reading, MA, second printing, 1969.
7. C. D. Olds, Continued Fractions, Random House, New York, 1963.
8. O. Perron, Die Lehre von den Kettenbrüchen, Band I, Teubner, Stuttgart, 1954.
9. W. Wythoff, A modification of the game of Nim, Nieuw Arch. Wisk., 7 (1907) 199–202.
10. A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, translated by J. McCawley, Jr., revised and edited by B. Gordon, vol. 2, Holden-Day, San Francisco, 1967.

MISCELLANEA

SONNET

76.

R. P. BOAS

Lines written after reading too many abstracts of talks at a Mathematics meeting (after Shakespeare, Sonnet 130, "My mistress' eyes are nothing like the sun")

No diagrams within my work commute;
Language will do. Against the tide of groups,
Lie, semisimple—I'm with King Canute.
Let others prate of posets and of loops,
Functors and morphisms, maximal ideals;
Give me the clichés of an earlier age.
Let no nonstandard models of the reals,
Sur- or bijections decorate my page.
The complex plane contains enough; for me
No sheaves of germs upon a manifold.
I'll never be approved by Bourbaki;
Words grow apace, but still my soul's not sold.
  And yet I think my work was as profound
  As this, tricked out with terms of modish sound.

CALCULUS IS ALGEBRA

WILLIAM S. HATCHER
Department of Mathematics, Laval University, Quebec, Canada G1K 7P4

1. Introduction. Presumably everyone who teaches calculus will agree that some knowledge of algebra is a prerequisite for a course in calculus. However, there can always be an argument as to how much is enough. The basic fact to be assimilated is that the real numbers form a complete ordered field, but this fact can be learned operationally without being assimilated conceptually (or at least without being totally and explicitly assimilated on the conceptual level). In any case, the basic problem is pedagogical and not mathematical and involves determining the level of algebraic conceptualization which optimizes the learning of calculus.

The main point of the present paper is that the use of a few extremely elementary notions and facts of ring theory enables one to do calculus in the hyperreal number system, thus reducing calculus totally to algebraic manipulation. In other words, with only a slight increase in conceptual sophistication, there is a substantial payoff in the form of reduced operational complexity. Since reducing operational complexity through conceptualization is the very heart of mathematical activity, I feel that this approach to the teaching of calculus lays serious claim to being the "correct one."

2. Algebraic prerequisites. To construct and use the hyperreal numbers in their simplest form, we need only the following notions and facts.

Notions.

(1) Ring with unit.
(2) Subring and ideal.
(3) Quotient ring by an ideal.
(4) Integral domain and field.
(5) Prime ideal and maximal ideal.

Facts.

(1) A maximal ideal is prime.
(2) Any proper ideal is contained in a maximal one.
(3) Every field is an integral domain.
(4) A quotient ring is a field iff the dividing ideal is maximal.

We will now see how these facts and notions are used to construct and then to use the hyperreal numbers. We take the real numbers R as given and understood on whatever level of conceptual sophistication is deemed appropriate by the instructor.

3. The construction of R^* from R. We begin by forming the ring R^N of all sequences of real numbers (N is the set of positive integers). Addition and multiplication of sequences is done in the obvious component-by-component fashion. The zero sequence is the zero of the ring and the sequence with constant value 1 is the unit of the ring R^N. In fact, R is embedded into R^N by the diagonal correspondence $r \mapsto \vec{r}$ where $\vec{r}$ is the sequence which is constantly r, $\vec{r}_n = r$ for all

The author is Professor of Mathematics at Laval University, Quebec, Canada. After earning Bachelor's and Master's degrees at Vanderbilt University, he received the degree of *Docteur ès Sciences* from the University of Neuchâtel, Switzerland, in 1963, with a thesis in mathematical logic. He has since published over thirty papers, monographs, and books in the fields of logic and foundations, algebra, computer science, and philosophy. He has lectured at major universities in Europe and North America, and spent the academic year 1972–73 as invited professor of computer science at the Swiss Federal Institute of Technology in Lausanne. In 1978 he was elected to a three-year term on the Council of the Association for Symbolic Logic. He maintains research activity in logic and in algebra and a special interest in the philosophy of science and religion. He is currently on the editorial board of *Baha'i Studies* to which he has also contributed as an author (see *The Science of Religion*, *Baha'i Studies*, vol. 2, rev. ed., 1980, as well as *Zygon/Journal of Religion and Science*, vol. 14, pp. 229–253, 1979). His *The Logical Foundations of Mathematics* is forthcoming with Pergamon Press, Oxford.

$n \in N$.

R^N is not a field, however, since it is not even an integral domain. For example, consider the two sequences s and t where s is 1 on the evens and 0 on the odds while t is 1 on the odds and 0 on the evens. Then $st = 0$, but neither s nor t is 0.

We want now to obtain a quotient ring R^N/M which is a field. We are thus looking for an appropriate maximal ideal M in R^N. We want the resulting field, called R^*, to satisfy (at least) the following conditions: (I) It must be totally ordered and must contain R as an ordered subfield (it must in fact preserve the embedding of R into R^N). (II) It must be non-Archimedean over R, i.e., it must contain elements ω which are bigger than every real number. (III) Every finitary operation $f\colon R^n \to R$, $n \geqslant 1$, and every finitary relation $K \subset R^n$, $n \geqslant 1$, must extend canonically to R^*. In each case, the extension defined on R^* must be identical on R with the original function or relation defined on the field R, and of the same arity. (IV) Moreover, for each $n \geqslant 1$, complements, unions, and characteristic functions of n-ary relations are preserved under canonical extensions. The composition of operations is preserved under the canonical extension of operations, and, for $n \geqslant 1$, the extension of an n-ary projection on R is the corresponding projection on R^*, and the extension of a constant function is constant. The field operations on R^* are the canonical extensions of the corresponding field operations on R.[†]

We now proceed to find an appropriate maximal ideal M.

DEFINITION 1. Let a be any sequence in R^N. By the *support* of a we mean the set $\sigma(a) \subset N$ of indices n such that $a_n \neq 0$.

Since the ideal becomes the zero element in the quotient ring obtained by dividing by the ideal in question, we are looking for an ideal of sequences having lots of zero values.

DEFINITION 2. F is the set of sequences a having a finite support set $\sigma(a)$.

Thus, the sequences in F are sequences which are zero everywhere except on some finite set of indices.

PROPOSITION 1. *F is a proper ideal in R^N.*

Proof. A simple, straightforward verification that the ideal properties are true of F.

Since F is a proper ideal, it is contained in a maximal ideal M.

DEFINITION 3. $R^* = R^N/M$. Elements of R^* are called *hyperreal numbers.*[‡]

[†] The conjunction of conditions I–IV is equivalent to the axioms of Keisler [7]. The main difference is that we here deal with n-ary relations and n-ary operations whereas Keisler deals rather with n-ary partial functions. Also, our approach here is totally "objective" in that we deal with functions and relations in extension in the way which is customary to the working mathematician, thereby avoiding any explicit appeal to logic or to such logical, syntactical notions as that of a formula. Other recent versions of nonstandard analysis (see [6]) involve far more logical notions than does Keisler and are therefore at the opposite end of the "objective-subjective" spectrum from the approach of this paper. The purpose of conditions like I–IV or Keisler's axioms is to obtain the so-called "transfer principle" which enables us to transfer wholesale from R to R^* all elementary statements (roughly, statements which quantify only over elements and not over subsets of R). This makes R^* an *elementary extension* of R (see Chang and Keisler [1] and Keisler [8]). There are some intentional redundancies in our formulation of conditions I-IV.

[‡] In the presence of the continuum hypothesis, the field $R^* = R^N/M$ is uniquely determined up to an isomorphism of ordered fields even though the maximal ideal $M \supset F$ is not necessarily unique (see [3]). It would therefore seem reasonable to tell students that R^* is uniquely defined for all practical purposes since models of set theory in which the axiom of choice (which we have assumed in the form of Krull's Lemma as fact (2) above) holds but CH fails are rarely, if ever, used in everyday mathematics. Also note that $R^* = R^N/M$ can be defined from R^N using ultrafilters on the powerset of the natural numbers N (see Keisler [8]). Indeed, the set S of Definition 4 and Proposition 2 of this paper is such an ultrafilter. This conversion between maximal ideals and ultrafilters is a special case of a natural bijection between filters on the powerset of N and ideals in R^N (see [2]). According to this bijection, the ideal F of Definition 2 above corresponds to the Fréchet filter, making ideals $M \supset F$ correspond to free ultrafilters. Thus, the R^* of Definition 3 above is an ultrapower of the reals R and a proper elementary extension of R. In view of the fact that the Boolean Prime Ideal theorem is somewhat weaker than the axiom of choice (see [4]), the construction using ultrafilters can be carried on under somewhat weaker ontological assumptions if this is considered desirable.

A hyperreal number is thus an equivalence class of sequences of real numbers. We use brackets [ ] to denote equivalence classes. Two sequences a and b are equivalent iff their difference $a - b \in M$. In particular, since $F \subset M$, two sequences are equivalent whenever they differ only on some finite set of indices. Also, every sequence $a \in M$ must have at least one zero value, $a_k = 0$ for some $k \in N$, for otherwise the sequence $1/a_n$ exists and the product $(1/a) \cdot a = \vec{1} \in M \Rightarrow M = R^N$, contradicting the maximality of the ideal M.

Since M is maximal, R^* is a field. Moreover, R is contained as a subfield of R^* by the obvious representation $r \mapsto [\vec{r}]$. The only thing to check is that this correspondence is injective, i.e., $[\vec{r}] = [\vec{s}] \Rightarrow r = s$. We have: $[\vec{r}] = [\vec{s}] \Leftrightarrow \vec{r} - \vec{s} = \overrightarrow{r - s} \in M$. Since the constant sequence $\overrightarrow{r - s}$ is either everywhere zero or nowhere zero, it must be constantly zero if it is in M. Thus $r = s$ if $[\vec{r}] = [\vec{s}]$.

It follows from the fact that $R \subset R^*$ that the cardinalities of R and R^* are the same (namely $2^{\aleph_0}$). Indeed, by definition the cardinality of R^* is less than or equal that of R while the above shows that the cardinality of R is less than or equal that of R^*.

We want now to extend the order relation $<$ on R to R^*. To do this smoothly, the following definition is useful.

DEFINITION 4. $X \in S \Leftrightarrow X = \sigma(a)$ for some $a \notin M$. S is the collection of all support sets of sequences not in M.

PROPOSITION 2. *The collection S satisfies the following properties*: (1) $N \in S$. (2) $X, Y \in S \Rightarrow X \cap Y \in S$. (3) $Y \supset X \in S \Rightarrow Y \in S$. (4) $\varnothing \notin S$. (5) *All cofinite sets (sets whose complements in N are finite) are in S.* (6) *For all $X \subset N$, either $X \in S$ or $(N - X) \in S$ and not both.* (7) *If $(X \cup Y) \in S$, then either $X \in S$ or $Y \in S$.* (8) *If $X \in S$, then X cannot be the support set of any sequence $a \in M$.*

Proof. (1) $N = \sigma(\vec{1})$ and $\vec{1} \notin M$. (2) If $X = \sigma(a)$ and $Y = \sigma(b)$ where $a, b \notin M$, then $X \cap Y = \sigma(a \cdot b)$. If $a \cdot b \in M$, then either $a \in M$ or $b \in M$ since M is maximal and therefore prime. Thus, $a \cdot b \notin M$ and $X \cap Y \in S$. (3) Suppose $Y \supset X \in S$, $X = \sigma(a)$ where $a \notin M$. Define the sequence b as follows:

$$b_n = \begin{cases} 1 & \text{if } n \in Y \\ 0 & \text{otherwise} \end{cases}.$$

Clearly, $Y = \sigma(b)$ and $a \cdot b = a$. By the absorbing property of the ideal M, $a = a \cdot b \in M$ if $b \in M$. Since $a \notin M$, $b \notin M$. Thus, $Y \in S$. (4) The null set $\varnothing$ is the support set of only one sequence, namely the constantly zero sequence $\vec{0}$. Since $\vec{0} \in M$, $\varnothing \notin S$. (5) Since $F \subset M$, all sequences with finite support are in M. If some sequence a with cofinite support is in M, then let b be any sequence with finite support $\sigma(b) = N - \sigma(a)$. The sequence $(a + b) \in M$ since M is closed under addition. But the sequence $a + b$ has no zero value (its support is N) and this is impossible as we have already seen. Thus, no sequence with cofinite support is in M and hence every cofinite set is in S. (6) Consider any $X \subset N$. Let the sequence a be the characteristic function of X and let a' be the characteristic function of $N - X$. Clearly, $X = \sigma(a)$ and $N - X = \sigma(a')$. If both $a \in M$ and $a' \in M$, then $(a + a') \in M$. But the sequence $a + a'$ has no zero value which is impossible for any sequence in M. Thus, either $a \notin M$ or $a' \notin M$ and consequently either $X \in S$ or $(N - X) \in S$ by definition. Finally, if both X and $N - X$ are in S, then by (2) above their intersection $\varnothing \in S$ contradicting (4). (7) Suppose that $(X \cup Y) \in S$ but $X \notin S$ and $Y \notin S$. Then, by the property (6) of S, $(N - X) \in S$ and $(N - Y) \in S$ whence their intersection is in X by property (2). By DeMorgan's law, this intersection is the complement in N of $X \cup Y$. We thus have $(N - (X \cup Y)) \in S$ and $(X \cup Y) \in S$ contradicting property (6). Thus, either X or Y must be in S. (8) Suppose $X = \sigma(a)$, $X = \sigma(b)$, $a \in M$, and $b \notin M$. Clearly $a \neq \vec{0}$ since otherwise $X = \varnothing \in S$, contradicting property (4) above. Define a new sequence c by the rule

$$c_n = \begin{cases} 1/a_n & \text{for } n \in X \\ 0 & \text{otherwise} \end{cases}.$$

By the absorbing property of the ideal M, the product sequence $c \cdot a \in M$. The sequence $c \cdot a$ has

the value 1 for every $n \in X$ and 0 elsewhere. Thus, the product $b \cdot (c \cdot a) = b$. But, again by the absorbing property of the ideal M, the product $b = b \cdot (c \cdot a) \in M$, contradicting the hypothesis that $b \notin M$. Thus, if $X \in S$ (i.e., if X is the support set of some sequence $b \notin M$), then X cannot be the support set of any sequence $a \in M$.

The usefulness of the collection S and its properties will be immediately clear from the ensuing development.

LEMMA 1. *For any $a, b \in R^N$, $[a] = [b]$ iff $\{n \mid a_n = b_n\} \in S$.*

Proof. $[a] = [b] \Leftrightarrow (a - b) \in M$. Now, $\{n \mid a_n \neq b_n\} = \{n \mid (a-b)_n \neq 0\} = \sigma(a - b)$. Thus, if $(a - b) \notin M$, then $\{n \mid a_n \neq b_n\} \in S$ and therefore $\{n \mid a_n = b_n\} \notin S$ by property (6) of S. This establishes the implication $\{n \mid a_n = b_n\} \in S \Rightarrow [a] = [b]$.

Going the other way, suppose $(a - b) \in M$ but $\{n \mid a_n = b_n\} \notin S$. Then, again by the property (6) of S, $S \ni \{n \mid a_n \neq b_n\} = \sigma(a - b)$ which contradicts property (8) of S. Thus, $[a] = [b] \Rightarrow \{n \mid a_n = b_n\} \in S$.

Lemma 1 gives us a useful necessary and sufficient condition for two sequences to be equivalent.

DEFINITION 5. $[a] < [b]$ iff $\{n \mid a_n < b_n\} \in S$, for $a, b \in R^N$.

By Lemma 1 and the properties of S, the relation $<$ is well-defined on equivalence classes of sequences.

THEOREM 1. *The relation $<$ is a total order on R^* which extends the usual total ordering on R.*

Proof. $\{n \mid a_n < a_n\} = \varnothing \notin S$. Thus $[a] \not< [a]$ for all $[a] \in R^*$. If $[a] < [b] < [c]$, then $\{n \mid a_n < c_n\} \supset \{n \mid a_n < b_n\} \cap \{n \mid b_n < c_n\}$ and this latter intersection is in S. Thus, by property (3) of S, the superset $\{n \mid a_n < c_n\}$ is in S and $[a] < [c]$ by definition.

These two facts establish that $<$ is a partial ordering of R^*. We now show that any two elements $[a]$ and $[b]$ of R^* are comparable. Suppose $[a] \not< [b]$, i.e., $\{n \mid a_n < b_n\} \notin S$. By property (6) of S, its complement $\{n \mid a_n \not< b_n\} = \{n \mid a_n \geqslant b_n\} \in S$. By definition, this last set $\{n \mid a_n \geqslant b_n\} = \{n \mid a_n > b_n\} \cup \{n \mid a_n = b_n\}$. We now have two sets whose union is in S. Thus, by property (7) of S, either $\{n \mid a_n > b_n\} \in S$ or $\{n \mid a_n = b_n\} \in S$, that is, by Definition 5 and Lemma 1, either $[a] > [b]$ or $[a] = [b]$.

It is not difficult to verify that $<$ is compatible with the field operations, making R^* an ordered field.

Finally, we observe that for diagonal sequences $\vec{r}$ and $\vec{s}$, $\{n \mid \vec{r}_n < \vec{s}_n\}$ is either the whole set N or else $\varnothing$ depending on whether $r < s$ holds or not. Since $N \in S$ and $\varnothing \notin S$, $[\vec{r}] < [\vec{s}] \Leftrightarrow r < s$. In other words the ordering $<$ on R^* extends the usual ordering on R. This completes the proof.

Theorem 1 shows that R^* satisfies condition I outlined above. We now turn our attention to the task of establishing property II, i.e., that R^* is non-Archimedean over R. We begin by establishing some useful terminology.

The absolute value function $|\ |: R^* \to (R^*)^+$ is defined in R^* in the usual way for ordered fields: For $a \in R^*$, $|a| = a$ if $a \geqslant 0$ and $-a$ otherwise. We also extend the ordering relation to sets of hyperreals in the usual way. This facilitates the following definition.

DEFINITION 6. A hyperreal number a such that $|a| > R^+$ is an *infinite* number. A hyperreal number a such that $|a| < R^+$ is an *infinitesimal* number. Finally, a *finite* number is a noninfinite one.

Clearly, 0 is the only real infinitesimal number. Moreover, the positive infinitesimal and the positive infinite numbers are in obvious 1–1 correspondence via inversion: $a > R^+$ if and only if $0 < 1/a < R^+$. This recalls the bijection between the real intervals $(0, 1)$ and $(1, \infty)$. A finite number is one whose absolute value is smaller than some positive real integer.

We now prove that there are infinite and therefore infinitesimal hyperreal numbers.

THEOREM 2. *There exists an element $\omega \in R^*$ which is greater than every real number $[\vec{r}] \in R^*$, $r \in R$.*

Proof. Let $s\colon N \to R$ be the identity sequence on N, i.e., $s_n = n$ for all $n \in N$. Let $\omega = [s]$. Let any real number $[\vec{r}]$ be given. Since R is Archimedean, there exists some natural number k such that $r < k$. Thus, $\{n \mid \vec{r}_n < s_n\} \supset \{n \mid n > k\}$. But this last set is in S since it is cofinite (it is the complement of the finite set $\{1,2,\ldots,k\}$). Thus, by property (3) of S, $\{n \mid \vec{r}_n < s_n\} \in S$ and $[\vec{r}] < \omega$. But $[\vec{r}]$ was arbitrary. Thus, ω is greater than every real number: $R^+ < \omega$.

The hyperreal number $1/\omega$ will be a positive infinitesimal number as we have already seen in the informal remarks preceding Theorem 2. Thus, the existence of infinite and of infinitesimal numbers is clearly established.

Notice that $r + \omega$ will also be infinite for any real r. In particular, for r negative this can be seen as follows: If $x = \omega - r$ were finite where $r \in R^+$, then $\omega = x + r$ is the sum of two finite numbers. But the finite numbers are easily seen to be closed under addition. Thus, $x = \omega - r$ has to be infinite. Hence the order structure of the positive infinite numbers contains at least the set $R + \omega$ which is order-isomorphic to the reals. In fact, the order structure of the positive infinite numbers is considerably more complicated as will be clear once the structure of the ring of finite numbers is examined further on.†

We conclude the present section by establishing III, namely that every function and relation on R has a canonical extension to R^*, and IV.

Given a real function $f\colon R \to R$, we define its canonical extension to R^* in the following manner: First, we extend f to R^N by the obvious rule: $f(s)_n = f(s_n)$ where $s\colon N \to R$ and $n \in N$. Finally, we extend f to equivalence classes $[s]$ of sequences by the definition $f([s]) = [f(s)]$. We must, however, justify that this is well-defined, i.e., that $[f(s)] = [f(t)]$ whenever we have $[s] = [t]$. By Lemma 1, this last condition is equivalent to the condition $\{n \mid s_n = t_n\} \in S$. Since $f(s_n) = f(t_n)$ whenever $s_n = t_n$, $\{n \mid f(s)_n = f(t)_n\} \supset \{n \mid s_n = t_n\}$ and so $\{n \mid f(s)_n = f(t_n)\}$ is in S by (3) of Proposition 2. Thus $[f(s)] = [f(t)]$ by Lemma 1. By the way we have extended functions f from domain R to domain R^N, it is clear that the extension of f will yield a diagonal sequence $\overrightarrow{f(r)} = f(\vec{r})$ when applied to a diagonal sequence $\vec{r}$. Since it is equivalence classes $[\vec{r}]$ that represent real numbers r as elements of R^*, the extension of f to R^* will yield the same corresponding values when applied to $[\vec{r}]$ as it did when applied to r to begin with, i.e., $f([\vec{r}]) = [\overrightarrow{f(r)}]$. Generalization to the case of a function $f\colon R^n \to R$, $n \in N$, of any nonzero finite number of variables is immediate. This establishes condition III for functions.

Condition IV is also immediate for functions since extension clearly preserves functional composition, projections, constant functions, and the field operations. We give the same names to functions on R and their canonical extensions to R^*.

Let us now examine the case of relations on R. The extension to R^* of a binary relation K on R is defined by: $[s]K[t]$ if and only if $\{n \mid s_n K t_n\} \in S$. This is well-defined on equivalence classes of sequences by Lemma 1 and the properties of the set S. Since $N \in S$ and $\varnothing \notin S$, it follows that K will hold between two equivalence classes of diagonal sequences, $[\vec{r}]$ and $[\vec{s}]$, when and only when K holds between r and s, i.e., rKs. Thus, the extension of K to R^* will continue to hold for exactly the same couples $\langle[\vec{r}],[\vec{s}]\rangle$ for which rKs on R. Since $N \in S$, the extension of $R \times R$ to R^* will be the universal relation $R^* \times R^*$. In fact, extension preserves complements, unions, and characteristic functions of binary relations, by the properties of S and the relevant definitions. We give the same names to relations on R and to their canonical extensions on R^*. Finally, the generalization to the case of n-ary relations, for $n \geqslant 1$, is immediate. This establishes conditions III and IV for relations and completes our task of constructing a field satisfying conditions I–IV.‡

Notice that the $<$ relation that we have previously defined on R^* (Definition 5) is the canonical extension of $<$ on R. Thus, the absolute value function $|\ |$ we have defined on R^* is the canonical extension of $|\ |$ defined on R. Also, the set N of natural numbers has a canonical

†See [5] for a study of the order structure of R^*.

‡One mildly subtle point remains to be checked: n-ary operations f are also $(n+1)$-ary functional relations. We need to know that the extension of such an f as a function is the same as the extension of f as a relation, in all cases. It is, and the verification is left to the reader.

extension $N^* \subset R^*$: $[s] \in N^* \Leftrightarrow \{n \mid s_n \in N\} \in S$. Elements of $N^* - N$ are infinite elements of R^* and are called *infinite natural numbers.*

The practical meaning of conditions III and IV is that all those identities and relationships which hold on R will continue to hold for the corresponding extensions defined on R^*. For example, the identity $\mathrm{Sin}^2 x + \mathrm{Cos}^2 x = 1$ for all real numbers x will hold, for all hyperreal numbers z, for the canonical extensions of Sin and Cos. The extension to R^* of the function f defined on R by the rule $f(x) = x^2$ will be defined by the rule of squaring when applied to any hyperreal number z. Etc.

Having now constructed R^*, we turn to a study of its structure.

4. The structure of R^*. An interesting aspect of the approach to the calculus via the hyperreal numbers is that we will henceforth have only occasional need for our construction of R^*. Now that we know that R^* satisfies the conjunction of I, II, III, and IV, we can "throw away" our construction of it. Of course, it is both pedagogically and theoretically useful to have a particular model of R^* at hand to help us think about the structure of the hyperreals.

This situation is analogous to the study of the real numbers in which a model of a complete ordered field can be constructed in several different ways, e.g., equivalence classes of Cauchy sequences or Dedekind cuts. The analogy is not perfect, however, since all models of complete ordered fields are isomorphic to each other, whereas there are nonisomorphic structures satisfying the conjunction of I–IV. Nevertheless, the analogy is close enough to permit a purely axiomatic treatment of calculus in the hyperreals as Keisler does in [7].

We let R_0 be the set of all finite hyperreal numbers. We have already noted in passing that

THEOREM 3. *R_0 is a subring of R^* but not a subfield.*

Proof. It is easy to check that the sum, difference, and product of finite numbers are finite. Infinitesimals are finite, and the reciprocals of nonzero infinitesimals are infinite, contradicting closure under division.

Let I denote the set of infinitesimal numbers. It is easy to check that I is an ideal in R_0. Thus, in particular, $I \cdot R_0 \subset I$. The equivalence classes $x + I$, $x \in R_0$, determined by I are called *monads*. Furthermore,

THEOREM 4. *I is a maximal ideal of R_0.*

Proof. Any ideal J that properly contains I must contain at least one finite, noninfinitesimal number x. But $1/x$ is also finite, noninfinitesimal and so $1 = (1/x) \cdot x \in J$ which yields $J = R_0$, i.e., J is not a proper ideal of R_0. This establishes the maximality of I.

As in our earlier construction, we again have a field, R_0/I. We establish that, in fact, $R \cong R_0/I$.

Let g: $R_0 \to R_0/I$ be the canonical homomorphism of R_0 onto the quotient field R_0/I. For all $x \in R_0$, $g(x) = x + I$. We show that there is exactly one real number r in each equivalence class $g(x) = x + I$. Uniqueness is easy, for if r_1 and r_2 are real numbers such that $r_1 - r_2 \in I$, then $r_1 - r_2 = 0$ since the difference of two real numbers is real and 0 is the only real number in the ideal I of infinitesimals.

Let, now, $x + I$ be some equivalence class for $x \in R_0$. We assume $x > 0$, the case $x \leqslant 0$ being handled symmetrically. If x is real, there is nothing to prove. Otherwise, let $X = \{r \in R \mid r \leqslant x\}$. The set X of real numbers is not empty since $0 \in X$. Furthermore, X is bounded above since x is finite and positive (in other words there are real numbers t such that $t > x \geqslant X$). X is thus a nonempty set of real numbers bounded above. By the completeness property of the field of real numbers R, X has a real supremum $s \geqslant X$. If $s > x$, then $s \in x + I$, for otherwise $|s - x| = s - x > r$ for some positive real r, which immediately gives $s > s - r > x \geqslant X$, and $s - r$ is a real bound of X smaller than s, a contradiction. On the other hand, if $s < x$, then for all positive real numbers r, $r + s > s \geqslant X$ which means that $r + s \notin X$. By the definition of X, then, $r + s > x$ for every positive real r. In other words, $r > x - s = |s - x|$ for every positive real r. Hence, by

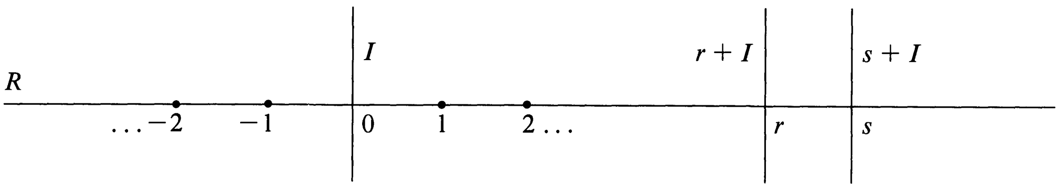
Definition 6, $s - x$ is infinitesimal which means $s \in x + I$. Thus, every monad $x + I$ is of the form $r + I$ for some unique real number r. We have proved:

THEOREM 5. *Every finite hyperreal number x is of the form $r + i$ where r is real and i is infinitesimal.*

We call r the *real part* and i the *infinitesimal part* of x. We define a mapping rp: $R_0 \to R$ which associates the real part of x to each finite hyperreal number x ("rp" stands for "real part"). The mapping rp is the composite of g: $R_0 \to R_0/I$ and the isomorphism between R_0/I and R and is therefore a ring homomorphism of R_0 onto R.

Notice that, for r, s real and i, j infinitesimal, $r + i < s + j$ if and only if $r < s$ or $r = s$ and $i < j$. We can therefore consider that the finite hyperreal numbers are ordered pairs of the form $\langle r, i \rangle$, r real and i infinitesimal, which are ordered lexicographically.

Let us now recall that R^*, and hence I, have the same cardinality as the real line R. We can thus use some fixed bijection between I and R to establish a bijection β between the (linearly) ordered ring R_0 and the real Cartesian plane $R \times R$. It is understood that this bijection is the identity on R (in other words, the abscissa is invariant under β^{-1}). We can use the bijection β: $R_0 \to R^2$ to induce on R^2 an ordered ring structure, making R^2 isomorphic to R_0 as an ordered ring. This yields the following geometric model of R_0:



Here, the abscissa is the usual real line and the ordinate is the ideal I of infinitesimals. At each point r of the real line the unique perpendicular at r represents the set $r + I$ which is the r-translation of the ideal I. It is the monad determined by r. Every finite hyperreal number is an ordered pair $\langle r, i \rangle$ whose real part is obtained by projecting onto the real axis and whose infinitesimal part is obtained by projecting onto the infinitesimal axis (the ordinate).

R_0 is an ordered ring, and two monads $r + I$ and $s + I$ are ordered as are their real parts. Geometrically, this order is given as increasing from left to right along the real line. In particular, the monad I of 0 is less than every positive real number.

It would be nice to be able to think of the order of I as given by the rule "greater means higher up," but the order-type of the infinitesimals does not permit this.[†] Even though the set I is bijective with the points on the real line, its order-type has no countable basis and cannot therefore be legitimately thought of as a sub-type of the order-type of the reals. Nevertheless, I is totally ordered and the translation operation $r + I$ preserves this order, so that each monad is ordered in the same way.

The positive infinite numbers can be thought of as "beyond" the real line in the positive direction while the negative infinites are beyond the real line in the negative direction. The positive infinite numbers are in 1-1 correspondence with the positive infinitesimals via inversion. A similar relationship holds for the negative infinite numbers.

Returning one last time to our geometrical model of R_0, let us imagine that we "compress" each vertical line $r + I$ into the point r. We now have simply the usual, one-dimensional geometrical model of the real line, except that now we think of the points on the real line as having a structure: each point on the abscissa *is* the monad $r + I$, which can be represented (as

[†]Except, of course, by analogy. Those who consider that the whole "real numbers = Euclidean line" concept is an analogy in the first place will not be restrained here. In any case, we can suppose a bijection between the positive infinitesimals and the portion of I above the real line, and this separation is preserved under translation.

usual) by the unique real number r it contains.† Since points are so small, their inner structure can be detected only by the "infinitesimal microscope" of [7].

The geometric model of R_0, both with and without the above modification, can be pedagogically useful in visualizing the structure of the hyperreals and in thinking about functions defined on the reals and their extensions to the hyperreals.

5. Using R^*. At this point, the development of the calculus using R^* is immediately accessible, without further theoretical development. A real function f is differentiable at a real point r iff, for any two nonzero infinitesimals i and j, the ratios $[f(r+i)-f(r)]/i$ and $[f(r+j)-f(r)]/j$ are in the same monad. In that case,

$$f'(r) = \mathrm{rp}\left(\frac{f(r+i)-f(r)}{i}\right)$$

where i is any nonzero infinitesimal.

The derivatives of polynomial functions can be simply calculated in short order and the usual differentiation formulas directly derived. The usual geometrical interpretations can be given and applications to maxima, minima, extrema, etc. speedily developed. Even the notion of a continuous function is not necessary to these developments, though it is easily accessible as well.

In these, and all subsequent developments, there is not only a practical simplification, but a theoretical one: the usual contravariant ε-δ limit definition is replaced by the equivalent covariant one involving infinitesimals. We say that $\mathrm{Lim}_{x\to a} f(x) = L$ if and only if $f(x) - L \in I$ whenever $x - a$ is, $x \neq a$. In particular, a continuous real function f is precisely one which preserves monads: $f\colon R \to R$ is continuous if, for every $r \in R$, $i \in I$, $f(r+i) = f(r) + j$ where $j \in I$. This is easily motivated and understood in terms of the geometrical model of R_0 above.

There is not only conceptual but deductive simplification as well. As an example, let us give a careful proof of the chain rule for the derivative of a composite function. This is a theorem whose proof in the classical ε-δ context is unnatural because of the possibility of a certain difference being zero.

THEOREM 6. *Let g be a real function differentiable at x while the real function f is differentiable at $g(x)$. Then the composite fg is differentiable at x and $(fg)'(x) = f'(g(x)) \cdot g'(x)$.*

Proof. Let $i \in I - \{0\}$ be given. For some infinitesimal k,

$$\frac{g(x+i)-g(x)}{i} = g'(x) + k.$$

Thus, the difference

$$j = g(x+i) - g(x) = (g'(x)+k)i$$

is infinitesimal. Similarly,

$$f(g(x)+j) - f(g(x)) = (f'(g(x)) + h)j$$

for some $h \in I$. Thus,

$$\frac{fg(x+i)-fg(x)}{i} = \frac{f(g(x)+j)-f(g(x))}{i} = \frac{(f'(g(x))+h)(g'(x)+k)i}{i}$$
$$= f'(g(x))g'(x) + t$$

where $t \in I$; $f'(g(x))g'(x)$ is thus the real part of the ratio $[fg(x+i)-fg(x)]/i$.

This is far from the only example of deductive simplification resulting from the systematic use of R^*.

†Let us remind ourselves that the points of the real line have a rich structure according to any of the well-known methods of constructing R from Q, e.g., Dedekind cuts or equivalence classes of Cauchy sequences.

6. Conclusions. The "modernization" of the teaching of calculus, implemented in the 1950's and 60's, involved a substantial increase in rigor as well as in linguistic and conceptual sophistication. It is not yet clear whether this really resulted in increased conceptual sophistication on the part of the student, and it almost certainly did not result in increased manipulative ability. I feel that using the hyperreal numbers as a framework for doing calculus holds forth the promise of producing both. The few basic concepts of ring theory necessary to a clear understanding of R^* represent real knowledge and not just an elaborate reformulation in set-theoretical terms of what is already known, as was unfortunately so often the case with some of the "modern" textbooks.

References

1. C. C. Chang and H. J. Keisler, Model Theory, North-Holland, Amsterdam, Netherlands, 1974.
2. P. C. Eklof, "Ultraproducts for algebraists" in Handbook of Mathematical Logic, J. Barwise, editor, North-Holland, Amsterdam, Netherlands, 1977, pp. 105–137.
3. P. Erdös, L. Gillman, and M. Henriksen, An isomorphism theorem for real-closed fields, Ann. of Math., 61 (1955) 542–554.
4. J. D. Halpern and A. Levy, "The Boolean prime ideal theorem does not imply the axiom of choice," in Axiomatic Set Theory, Amer. Math. Soc. Proc. Symposia Pure Math., vol. 1, 13, 1971, pp. 83–134.
5. W. S. Hatcher and C. Laflamme, On the order structure of the hyperreal line, Prepublication # MPA 32, Département de Mathématiques, Université Laval, 1981, 13 pages (to appear in Z. Math. Logik Grundlag. Math., 1983).
6. J. M. Henle and E. M. Kleinberg, Infinitesimal Calculus, M.I.T. Press, Cambridge, Mass., 1979.
7. H. J. Keisler, Elementary Calculus, Prindle, Weber & Schmidt, Boston, 1976.
8. ______, Foundations of Infinitesimal Calculus, Prindle, Weber & Schmidt, Boston, 1976.

ANY NEW HELLY NUMBERS?

EDWIN O. BUCHMAN

Department of Mathematics, California State University, Fullerton, CA 92634

FREDERICK A. VALENTINE

Department of Mathematics, University of California, Los Angeles, CA 90024

We often ask our friends "Any new Helly numbers?" Naturally, we mean "Have you found any new theorems of Helly type relating to families of convex sets?" Since Helly's theorem [5] states that "a family of bounded closed convex sets in Euclidean space E^n has nonempty intersection if and only if every $n + 1$ or fewer members of the family have nonempty intersection," then $n + 1$ is, by definition, a Helly number. It has been our gratifying experience that a casual inspection of a combinatorial situation often reveals some Helly numbers. Also, occasionally a trivial consequence of Helly's theorem immediately suggests a problem which one is unable to solve.

1. For instance, consider a convex polyhedron P in E^n.

Professor Buchman received his Ph.D. from the University of California, Los Angeles, in 1968 under Frederick Valentine. He then joined the department of mathematics at California State University, Fullerton. He has written several papers on convex sets and related fields of geometry, and his current research interests include the theory of geometric constructibility.

Professor Valentine earned his University of Chicago Ph.D. under G. A. Bliss and L. M. Graves. His thesis was on the calculus of variations. His main research in convexity has resulted in many articles and his book *Convex Sets*, McGraw-Hill, 1964, reprinted by Krieger, 1976. He held an instructorship at the University of Tennessee before going to UCLA where he is now an Emeritus Professor. He spent sabbaticals at the University of Chicago, the University of Washington, the University of Auckland, New Zealand, and twice in the Republic of South Africa.

THEOREM 1. *Suppose for each set of $n+1$ faces of P there exists at least one point $p \in P$ from which one can drop $n+1$ perpendicular segments respectively to these given faces (the point p may vary with the choice of the $n+1$ faces). Then there exists at least one point $q \in P$ such that for each face F of P a perpendicular segment from q to F exists.*

This theorem is a direct consequence of the fact that every $n+1$ of the vertical strips to the faces of P have at least one point in common, and Helly's theorem [5] implies the result. (A vertical strip to a face F is the union of all lines intersecting F which are perpendicular to F.) Now comes the interesting part.

THEOREM 1A. *In the plane E^2 if a convex polygon has at most 4 edges, then the Helly number $h = n+1$ in the preceding Theorem 1 can be reduced to $h = 2$ instead of $h = 2+1 = 3$.*

However, in E^n for $n > 2$ if the number of faces of P is less than $n+3$, we suspect that $h \leqslant n$, and perhaps $h = n$. We do not know how to prove this even for E^3. The proof of Theorem 1A is given at the end of this article, so as not to interfere with the natural flow of ideas.

Note. In Theorem 1 for the plane E^2 if the polygon P has at least 5 edges, then the Helly number $h = 3$ is best, as illustrated by Fig. 1. Observe that for each pair of edges perpendiculars can be dropped to them from some common point in P, yet no such point exists for all five edges.

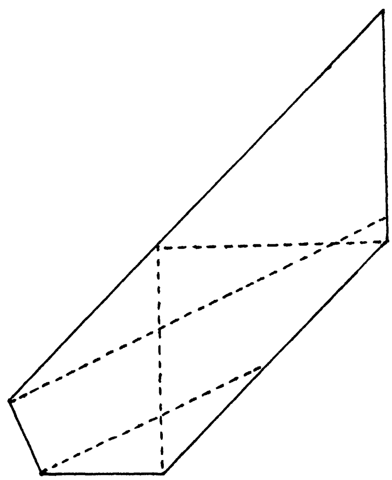


FIG. 1

The above theorem suggests an even more interesting set of problems for convex polyhedra in E^n, since we only have proofs for sets in the plane E^2.

THEOREM 2. *Suppose every $n+1$ points on the boundary of a convex polyhedron P in E^n can be covered by a vertical strip to some face of P. Then all of P can be covered by a vertical strip to some face of P. (Proved only for $n = 2$.)*

As before a vertical strip to a face F of P is the union of all lines intersecting F which are perpendicular to that face. In E^2 a vertical strip is sometimes called a vertical edge strip for emphasis. The proof of Theorem 2 for $n = 2$ is more difficult than that for Theorem 1, and it is given at the end of this article. Furthermore, we do not know how to prove it for $n > 2$.

Note. The Helly number $h = 3$ for $n = 2$ in Theorem 2 is best, since the following example (Fig. 2) illustrates a pentagon P in E^2 where every two points of the boundary of P can be covered by a vertical edge strip, yet no vertical edge strip covers all of P.

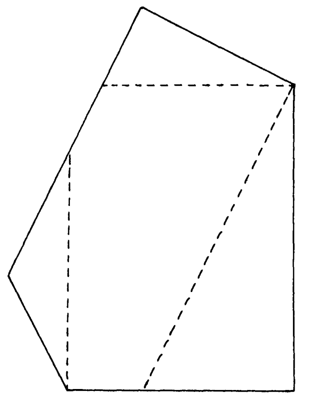


FIG. 2

2. There is a nice consequence of a theorem of Klee's which leads to a number of unanswered questions. Klee's theorem [**3**] states

> *If every $n+1$ or fewer members of a family $\mathfrak{F}$ of bounded closed convex sets in E^n are simultaneously contained in some translate of a convex set K, then all the members of $\mathfrak{F}$ are simultaneously contained in a translate of K.*

This result is also valid if the words "are contained in" are replaced by "contain" or "intersects."

Now consider the following.

THEOREM 3. *If every $n+1$ or fewer members of a family $\mathfrak{F}$ of bounded closed convex sets in E^n are simultaneously contained in some set of constant width d, then all the members of $\mathfrak{F}$ are simultaneously contained in a set of constant width d.*

Since the proof of this result is pleasing, and since it gives rise to other questions, it is sketched here in fairly complete form.

Proof. First observe that since every $n+1$ or fewer members of $\mathfrak{F}$ are contained in a set of constant width d, then for each direction R these $n+1$ sets are contained in a "strip" of width d whose bounding hyperplanes are perpendicular to R. Klee's theorem above implies that all the members of $\mathfrak{F}$ are contained in a strip of width d whose bounding hyperplanes are perpendicular to R. Since this holds for all directions R, all the members of $\mathfrak{F}$ are contained in a convex set C of diameter less than or equal to d. (The diameter of a set is the maximum of the distances between pairs of points of the set.) Since the set C of diameter d can be enclosed in a set of constant width d, Meissner [**1**], [**4**], Theorem 3 follows as a consequence.

Theorem 3 fails for containing sets of "width d" instead of "constant width d". For example, let $\mathfrak{F}$ be the four longest diagonals of a regular octagon.

Theorem 3 naturally suggests several other questions. For instance, "Suppose every 3 members of a family $\mathfrak{F}$ of bounded closed convex sets in the plane E^2 are contained in a Reuleaux triangle of width d (the family contains at least 3 members). What additional assumptions, if any, does one need so that all the members of $\mathfrak{F}$ are contained in a Reuleaux triangle of width d?"

By Theorem 3 we know that all the members of $\mathfrak{F}$ are contained in a set of constant width d, but it is not necessarily a Reuleaux triangle.

Using the word "contain" instead of "are contained in" Theorem 3 has the form

THEOREM 4. *If every $n+1$ or fewer members of a family $\mathfrak{F}$ of bounded closed convex sets in E^n simultaneously contain a set of width d, then all the members of $\mathfrak{F}$ simultaneously contain a set of width D, with $D \geqslant d$. (The width of a set S is the width of the narrowest strip which contains S.)*

To prove this let K be a line segment of length d. Since every $n+1$ or fewer members of $\mathfrak{F}$ contain a set of width d simultaneously, these members then simultaneously contain a translate of

the segment K [1], [5]. Since this holds for each direction R, Klee's theorem above implies that all the members of $\mathfrak{F}$ simultaneously contain a line segment $K(R)$ of length d in the direction R. The convex hull C of the union of the segments $K(R)$ for all directions R is a convex set of width D, where $D \geqslant d$.

As our good friend Ernst Straus pointed out, the set C of width $D \geqslant d$ cannot always be whittled down to a set C' of constant width d. For instance the equilateral triangle of width 1 does not contain a set of constant width 1. That raises the interesting question.

Question: What hypotheses, if any, should be added to Theorem 4, so that all the members of $\mathfrak{F}$ simultaneously contain a set of *constant* width d?

The most difficult questions are those dealing with "intersections." For instance,

"Suppose every $n+1$ members of a family $\mathfrak{F}$ of bounded closed convex sets in E^n intersect a set of constant width d simultaneously. What additional hypotheses, if any, does one need so that all the members of $\mathfrak{F}$ intersect a set of constant width d simultaneously?"

3. In conclusion, we sketch in fairly complete form proofs for Theorem 1A and Theorem 2 ($n = 2$).

Proof of Theorem 1A. Suppose that P is a convex quadrilateral with consecutive vertices a, b, c and d. To prove Theorem 1A, we will show that for each set of 3 edges of P there exists at least one point in P from which perpendiculars to these 3 edges can be dropped, and Theorem 1 then implies the result. So let ab, bc and cd be three consecutive edges of P. The vertical strips to ab and cd must intersect in P by hypothesis. To see that a point $p \in P$ in the intersection of these two strips exists from which perpendicular segments to ab, bc and cd exist, consider the three cases as shown in Fig. 3, namely where the angles at b and c are (i) both obtuse, (ii) both nonobtuse, (iii) one obtuse and one nonobtuse. The proof when P is a triangle is trivial. This completes the proof of Theorem 1A. A proof for E^3 would be most welcome.

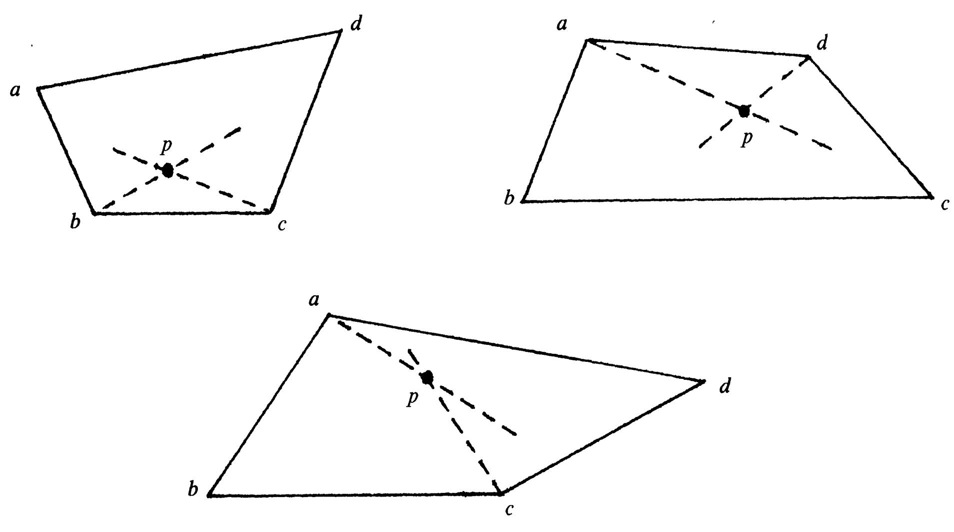


FIG. 3

Proof of Theorem 2 ($n = 2$). This proof is more involved, since we find it necessary to divide it into more cases as follows.

Case 1. Suppose P contains three nonobtuse angles at vertices a, b, c in its boundary. If P is a rectangle, then the conclusion of the theorem is immediate. Hence, assume P is not a rectangle. In this case all of the remaining angles of P are obtuse. See Fig. 4A. If ab, bc or ca is an edge of P,

then P is covered by a vertical strip to one of these edges. Hence, suppose such is not the case. Since the angles at a, b and c are nonobtuse, and since ab, bc and ca are not edges of P, the perpendicular lines to ac and bc at a and b respectively meet at a point d outside P such that angle $adb > 90°$. Similarly there exist points e and f such that angle $bec > 90°$ and angle $afc > 90°$. These three inequalities for the angles adb, bec and cfa imply that no segment in the boundary of P has a vertical strip which covers a, b and c, a contradiction.

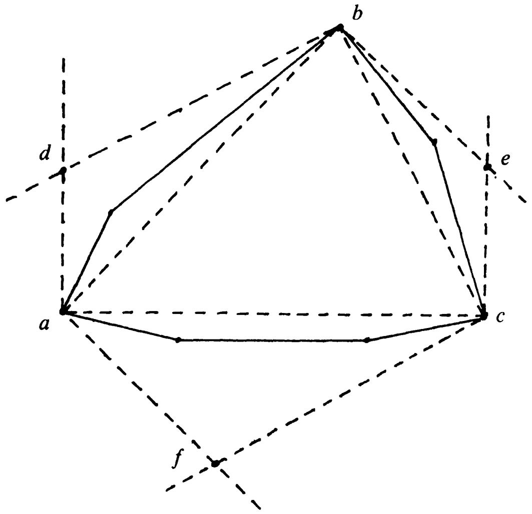


FIG. 4A

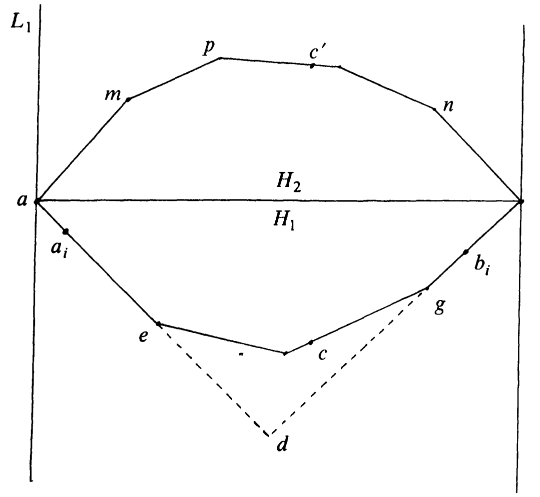


FIG. 4B

Case 2. Suppose P contains exactly two nonobtuse angles occurring at vertices a and b respectively. In such a situation lines of support L_1 and L_2 to P at a and b respectively exist which are both perpendicular to ab. If ab is an edge of P, then since angle $a \leqslant 90°$, angle $b \leqslant 90°$, the vertical strip to ab covers P. Hence, suppose ab is not an edge of P. Since angle $a \leqslant 90°$, angle $b \leqslant 90°$, there exist edges ae and bg of P such that angle eab + angle $gba \leqslant 90°$, and such that ae and bg lie in the same half-space H_1 determined by ab as in Fig. 4B. Also the extensions of ae and bg in H_1 must intersect in a point $d \in H_1$, as illustrated. The above implies angle $adb \geqslant 90°$.

If angle $adb = 90°$, $d \in P$, then the vertical strip to ad covers P since angle $a \leqslant 90°$. Hence, suppose angle $adb = 90°$ with $d \notin P$ or angle $adb > 90°$. (It should be realized that the points e, g and d need not be distinct.) This implies that no vertical strip to an edge of P in H_1 can cover points a_i and b_i chosen sufficiently close to a and b on ae and bg respectively. Let H_2 be the half-plane opposite to H_1 and let am and bn be the two edges of P which abut ab at a and b respectively, and which lie in H_2. Since all the angles in P except those at a and b are obtuse, one can choose a boundary point c' of P in H_2 such that neither am nor bn has a vertical strip which covers a_i, b_i and c' simultaneously. Thus the only vertical strip of P which covers a_i, b_i and c' has as its corresponding edge rs, which is interior to H_2 and strictly between L_1 and L_2. But the closed vertical strip to any such edge rs cannot cover both a and b, and hence cannot cover both a_i and b_i. Hence Theorem 2 is true in this case.

Case 3. Suppose P contains at most one nonobtuse angle. When such a vertex exists, it will be denoted by p. Use Fig. 4B assuming angle p is nonobtuse. There exist parallel lines of support to P, say L_1 and L_2, which miss p (when p exists). Let a and b be vertices of P which lie in L_1 and L_2 respectively. The segment ab is not an edge of P, since P has at most one nonobtuse angle at p (if it exists). In Fig. 4B, without loss of generality, suppose $p \in H_2$ and choose boundary points a_i and b_i in H_1 near a and b respectively. Since all the angles of P in H_1 and also the angles at a and b are obtuse, no vertical strip for an edge of P in H_1 can cover both a_i and b_i and also a point

$c \in H_1$ chosen appropriately. One choice for c is illustrated in Fig. 4B. Also, since the angles at a and b of P are both obtuse, the vertical strip to am or bn in Fig. 4B cannot cover both a_i and b_i. Hence, as in Case 2, a vertical strip to an edge rs interior to H_2 would have to cover a_i, b_i, and c. The rest of the argument is the same as in Case 2.

This completes the proof of Theorem 2 when $n = 2$. In order to obtain a proof for $n \geqslant 3$, it appears likely that one should first seek a simpler proof for $n = 2$. We tried to do this by induction, but our efforts failed. Hopefully some reader may discover a simpler solution. Good hunting!

Note: If the polygon P in Theorem 2 ($n = 2$) has at most 4 edges, the Helly number is $h = 2$. To see this, if all the angles of a quadrilateral are nonacute, it must be a rectangle and hence covered by a vertical edge strip. Hence, suppose the angle a in a convex quadrilateral $abcd$ is acute. If either angle b or angle d is nonobtuse, then either an edge strip to ab or to ad covers P. If both angle b and angle d are obtuse, the vertices a and c cannot be covered by a vertical edge strip, and the hypotheses are not satisfied. If P is a triangle, the conclusion is trivial.

The following modification of Theorem 2 is suggested.

Question. Does Theorem 2 still hold if "vertical strips" are replaced by "parallel strips"? A parallel strip to a face F is the union of all lines which intersect F and which are all parallel to some fixed line.

4. For results on Helly's theorem prior to 1963 the reader is encouraged to read the excellent paper "Helly's theorem and its relatives" by Danzer, Grünbaum and Klee [**2**]. A compilation of subsequent results would be most welcome, since combinatorial problems in geometry have attracted mathematicians in recent years.

References

1. T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Springer-Verlag OHG, Berlin, 1934, reprint Chelsea, New York, 1948.
2. L. Danzer, B. Grünbaum and V. Klee, Helly's Theorem and Its Relatives, Am. Math. Soc. Symp. on Convexity, Seattle, Proc. Symp. Pure Math., vol. 7, 1963.
3. Victor Klee, The critical set of a convex body, Amer. J. Math., 75 (1953) 178–188.
4. E. Meissner, Über Punktmengen konstanter Breite, Vjschr. naturforsch. Ges. Zürich, vol. 56, 1911, pp. 42–50.
5. F. A. Valentine, Convex Sets, McGraw-Hill, New York, 1964. Reprinted Krieger, Huntington, NY, 1976.

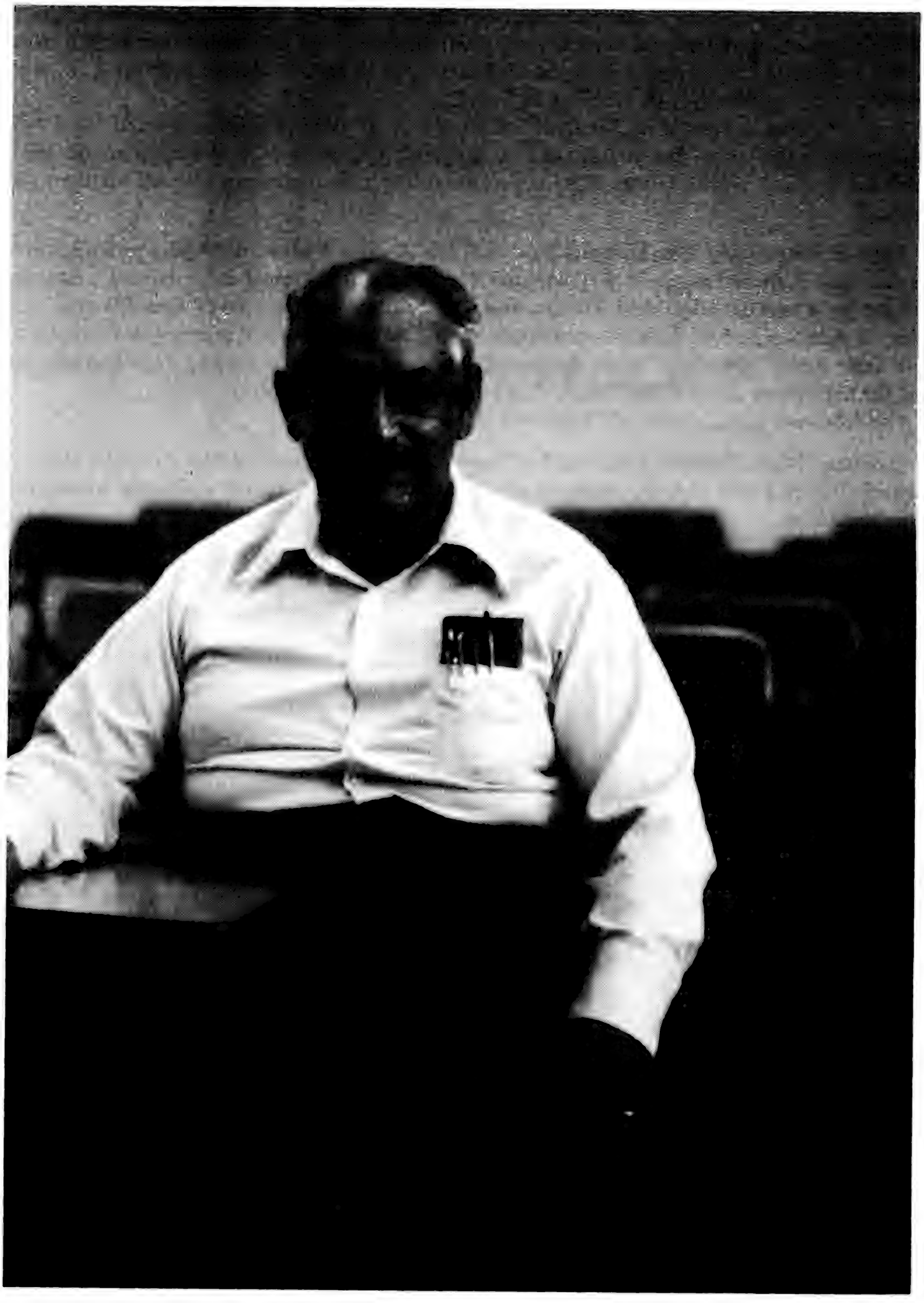
MISCELLANEA

77.

Leonid Hambro, the well-known pianist, told me recently that he was about to enter a billiards tournament in which he would play 12 games; he knew the opposition, he said, and he estimated his odds for winning any particular game as 8 to 5. "What do you think your chances are of sweeping all 12 games?" I asked him. "They're pretty small," he said. "The probability that I'll win any one game is 8/13. To find the probability that I'll win all 12 you have to take 8/13 to the 12th power. That's a pretty small number."

He did not have a calculator in his pocket. But he had a pencil and a pad—and an inspiration. "Hey!" he said. "Those are Fibonacci numbers. The ratio of successive terms approaches a limit (about .618), and very fast: even a ratio near the beginning like 8/13 is very close to the limit." He scribbled some additions. "The 12th Fibonacci number after 8 is 2584. Therefore 8/13 to the 12th power is approximately the same as 8/13 times 13/21 and so on, twelve times; everything cancels out except the 8 in the beginning and the 2584 at the end. So the probability that I will win all 12 games is about 8/2584, or about 1/300. See, I told you it was pretty small."

—LEONARD GILLMAN, The University of Texas at Austin



Few of us have our names become mathematical household words in our lifetime; this man did. Students have been learning his name for so long that most people would be surprised to hear that he lived till 1981. He published the theorem that made him famous in 1940, when he was twenty-six years old. The secret is revealed on p. 410.

unnecessary to determine the conjugates K_i completely. Unfortunately, determination of $\phi(\alpha)$ for α not in S_5 involves some calculations.

LEMMA 6. *The outer automorphism ϕ has the following values*:

$$\phi((12)) = (12)(36)(45)$$
$$\phi((13)) = (16)(24)(35)$$
$$\phi((14)) = (13)(25)(46)$$
$$\phi((15)) = (15)(26)(34)$$
$$\phi((16)) = (14)(23)(56)$$

Proof. The value of $\phi((1i))$ for $2 \leqslant i \leqslant 5$ is found, using Lemma 6, by determining $(1i)P_j(1i)$ for each j. This short computation is left to the reader.

The determination of $\phi((16))$ may be carried out as follows: Make a list of generators $\sigma_1, \dots, \sigma_6$ of the six Sylow 5-subgroups of S_5. Then $\rho(\sigma_1), \dots, \rho(\sigma_6)$ generate the six Sylow 5-subgroups of K_6. Next, conjugate these by (45) and by the powers of σ to obtain generators of the Sylow 5-subgroups in each K_i. Finally, select some K_i, some generator β of one of its Sylow 5-subgroups, form $(16)\beta(16)$, and locate the (unique) K_j in which it lies. Then $\phi((16))$ carries i to j. The result of this work is stated in the lemma. □

Since every element of S_6 is a product of transpositions of the form $(1i)$, the information above allows one to evaluate $\phi(\beta)$ for every $\beta \in S_6$.

COROLLARY 7. *If ϕ is an outer automorphism of S_6, then ϕ takes a* 3-*cycle to a product of two disjoint* 3-*cycles, but ϕ preserves the cycle structure of any permutation of order* 4.

Proof. Just evaluate ϕ on (123) = (13)(12) and on (1234) = (14)(13)(12). □

COROLLARY 8. *If ϕ is the automorphism of Theorem* 3, *then*

$$\phi^2 = \gamma_{\sigma^{-2}}, \quad \text{conjugation by} \quad \sigma^{-2}.$$

Proof. Using Lemma 7, one may show that $\phi^2(\beta) = \sigma^{-2}\beta\sigma^2$ whenever $\beta = (1i)$ and $2 \leqslant i \leqslant 6$; it follows that this formula holds for every $\beta \in S_6$. □

THEOREM 9. $\mathrm{Aut}(S_6) = \mathrm{Inn}(S_6)\langle\psi\rangle$ *with ψ an outer automorphism of order* 2. *Therefore*, $\mathrm{Aut}(S_6)$ *is a group of order* 1440 *that is a semidirect product of S_6 by a group of order* 2.

Proof. Define $\psi = \gamma_\sigma\phi$. It suffices to show ψ has order 2. If $\beta \in S_6$, then

$$\psi^2(\beta) = \gamma_\sigma\phi\gamma_\sigma\phi(\beta) = \gamma_\sigma\phi(\sigma\phi(\beta)\sigma^{-1}) = \gamma_\sigma(\phi(\sigma)\phi^2(\beta)\phi(\sigma^{-1})).$$

Since $\phi(\sigma) = \sigma$ (as we saw before Lemma 6), Corollary 8 shows these equations conclude with $\psi^2(\beta) = \beta$. □

References

1. H. A. Bender, A new method for the determination of the group of isomorphisms of the symmetric group of degree n, this MONTHLY, 31(1924) 287–289.

2. O. Hölder, Bildung zusammengesetzter Gruppen, Math. Ann., 46(1895) 321–422.

3. D. W. Miller, On a theorem of Hölder, this MONTHLY, 65(1958) 252–254.

4. J. J. Rotman, The Theory of Groups, An Introduction, 2nd ed., Allyn & Bacon, Boston, MA, 1973.

5. E. Witt, Die 5-fach transitiven Gruppen von Mathieu, Abh. Math. Sem. Univ. Hamburg, 12(1938) 256–264.

ANSWER TO "PHOTO" ON PAGE 376

Leonidas Alaoglu.

RADON INVERSION—VARIATIONS ON A THEME

ROBERT S. STRICHARTZ
Mathematics Department, Cornell University, Ithaca, NY 14853

1. Introduction. In 1917 Radon published **[8]** a solution to the problem: recover a function on the plane from its integrals over all lines in the plane. (Functions may be real or complex valued. Radon's paper is reproduced in Helgason's book **[5]**.) This was in fact a variation on a theme of Funk **[2]**, who in 1916 showed how to recover a function on the sphere from its integrals over all great circles. But mathematical terminology has been unkind to Funk, and the name *generalized Radon transform* has been applied to any operator that associates to a function on some geometric space its integrals over some class of geometric objects. The generalized Radon inversion problem is then to recover the function—if possible—from its generalized Radon transform. Perhaps the most direct generalization of Radon's original problem is to consider functions defined on an n-dimensional space and integrals over all k-dimensional affine subspaces, the so-called *k-plane transform*. Special cases of this are $k = n - 1$, called the *Radon transform*, and $k = 1$, called the *X-ray transform* because the exposure at a point on an X-ray picture gives a rough measurement of the integral of the function representing the density of the object being X-rayed along a line. If the X-rays form a parallel beam perpendicular to the photographic plate, then a single picture represents the X-ray transform evaluated on a family of parallel lines. In current practice most X-ray machines produce divergent (or fan) beams, so a single picture represents the X-ray transform evaluated on a family of lines passing through a fixed point. See **[4]** for a discussion of the mathematical problems that this entails. In either mode, since only a finite number of pictures are taken, the entire X-ray transform is not given. Thus a theoretical inversion of the X-ray transform is not equivalent to the problem of reconstructing an object from a finite set of X-rays. These issues are discussed in detail in the survey articles **[9]** and **[10]**. Other examples of generalized Radon transforms are discussed in Helgason's book **[5]**, and in numerous research articles, many of which can be found in the bibliography of **[5]**. Related mathematical problems are discussed in Zalcman **[12]**.

Generalized Radon transforms have been studied because they have practical applications, because they are of fundamental theoretical significance, or sometimes just because they lead to elegant mathematical problems. It is this last aspect that I want to emphasize in this article, which concludes with a collection of problems and solutions involving the inversion of various generalized Radon transforms. I hope the problems are attractive enough to tempt the reader to take pen in hand and attempt a solution. The problems can be solved by fairly elementary means, the solutions are of a conceptual rather than a technical nature, and in my opinion they are a lot of fun!

Before presenting the problems, I will discuss some solutions to Radon's original problem, since these will reveal some strategies that might prove useful in the sequel. I will also give an especially simple proof of the hole theorem and discuss a class of examples.

Autobiographical sketch of the author prepared at the request of the editors.

I graduated from the Bronx High School of Science in 1960, received a B. A. from Dartmouth College in 1963 and a Ph.D. in mathematics from Princeton University (under E. M. Stein) in 1966. I spent a year in France as a NATO Postdoctoral Fellow, then went to M.I.T. as a C. L. E. Moore Instructor, and in 1969 came to Cornell where I am now Professor of Mathematics. My major area of interest is harmonic analysis and its applications to many areas of mathematics. I have been fortunate to have many fine teachers, including Henrietta Mazen, Richard Williamson, Leon Henkin, Mischa Cotlar, A. Besicovitch, Eli Stein, S. Bochner, Harry Furstenberg and Irving Segal, from whom I learned not just the stuff of mathematics, but something of its spirit. What I love most about mathematics is the joy of discovery, when understanding overcomes confusion. What I like least about mathematics is the way clear and simple ideas tend to become muddy and murky and mystifying when committed to the printed page.

2. Radon Inversion. We begin by choosing a convenient parametric representation of all lines in the plane. Actually, if we want a one-to-one correspondence between parameters and lines, then there is no convenient representation. (This is essentially a consequence of the fact that the lines in the plane have a natural structure of a nontrivial vector bundle over the circle.) However, it is convenient enough to consider each line as given by the equation $x \cdot w = s$, where $w = (w_1, w_2)$ is a unit vector in the plane and s a real number. Using the pair (w, s) as parameters for the line gives a two-to-one correspondence, since the parameters $(-w, -s)$ correspond to the same line. Note that w is just a unit normal to the line ($w = (-\sin\theta, \cos\theta)$ if the line makes an angle θ with the x-axis) and $|s|$ is the distance between the line and the origin. The line $x \cdot w = s$ has a natural parametric representation (with parameter $t \in \mathbb{R}$)

$$x_1(t) = sw_1 + tw_2, \quad x_2(t) = sw_2 - tw_1,$$

and we can write the Radon transform $Rf(w, s) = \int_{x \cdot w = s} f = \int_{-\infty}^{\infty} f(x(t))\, dt$ for suitable functions f defined on the plane.

The quickest way to invert the Radon transform is to note the connection with the Fourier transform $\hat{f}(\xi) = \int_{\mathbb{R}^2} f(x) e^{-ix\cdot\xi}\, dx$. If we write ξ in polar coordinates $\xi = |\xi| w$, so $x \cdot \xi = |\xi| x \cdot w$, then $e^{-ix\cdot\xi}$ will be constant on all the lines $x \cdot w = s$. Thus in performing the x-integration in the Fourier transform we want to integrate first along these lines, giving $e^{-is|\xi|} Rf(w, s)$ and then in the perpendicular direction, $\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-is|\xi|} Rf(w, s)\, ds$. To paraphrase: *the two-dimensional Fourier transform is a one-dimensional Fourier transform of the Radon transform*. The two-dimensional Fourier inversion formula $f(x) = (2\pi)^{-2} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi$ then gives a Radon inversion formula,

$$f(x) = (2\pi)^{-2} \int_{\mathbb{R}^2} \left(\int_{-\infty}^{\infty} e^{-is|\xi|} Rf\left(\frac{\xi}{|\xi|}, s \right) e^{ix\cdot\xi}\, d\xi. \tag{2.1}$$

By expressing the ξ-integration in polar coordinates this simplifies slightly to

$$f(x) = \int_0^{\infty} \int_{S^1} \left(\int_{-\infty}^{\infty} e^{ir(x\cdot w - s)} Rf(s, w)\, ds \right) r\, dw\, dr. \tag{2.2}$$

Of course we must place hypotheses on f so that the Fourier inversion formula is valid in order to obtain this result.

This first inversion formula is not particularly illuminating since it involves a multiple integral that is not absolutely convergent. (The oscillations of the complex exponential give the s-integral sufficiently rapid decay in r to make the r-integral converge.) We might be tempted to tinker with it further, formally interchanging the orders of integration and trying to make sense out of the divergent expression that results. However, this approach leads to technicalities that can be avoided by another, more direct method. This second method involves looking at the dual Radon transform: Instead of integrating functions defined on points over all points lying on a fixed line, we integrate functions defined on lines over all lines containing a fixed point. Thus let $g(w, s)$ be a suitable function of lines. (Take g to be even; $g(-w, -s) = g(w, s)$, so it is truly defined on the lines $x \cdot w = s$ and not merely on the parameters (w, s) representing the lines.) For a fixed point z, all the lines passing through z have the form $x \cdot w = s$ where $s = z \cdot w$, so that the parameters $(w, z \cdot w)$ describe all these lines as w varies over the circle. Since the circle naturally parametrizes the set of lines we want to integrate over, there is no doubt that we want to define

$$R^*g(z) = \int_{S^1} g(w, z \cdot w)\, dw$$

(or more explicitly

$$R^*g(x, y) = \int_0^{2\pi} g\left(\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, x\cos\theta + y\sin\theta \right) d\theta)$$

for the dual Radon transform. The terminology is well justified because R^* is the adjoint operator of R, in the sense that

$$\int_{-\infty}^{\infty}\int_{S^1} Rf(w,s)g(w,s)\,dw\,ds=\int_{R^2} f(x)R^*g(x)\,dx$$

as can easily be verified by substituting the definition of R^*g in the right side and performing the x-integration first along the lines $x\cdot w=s$. We will not make use of this fact here; rather we want to emphasize that the composition operator R^*Rf makes sense from a geometric point of view as a first step in inverting R. The reason for this is that the value of f at a fixed point z only influences Rf along those lines passing through z, so the integral of Rf over these lines, $R^*Rf(z)$, is the simplest object we can construct from Rf that might have a connection with $f(z)$.

With this as motivation, let's compute what $R^*Rf(z)$ actually is. First let $z=(0,0)$. The lines through the origin are just $x\cdot w=0$, and if we let $w=(-\sin\theta,\cos\theta)$, then the line is given parametrically as $x(t)=t(\cos\theta,\sin\theta)$. Thus

$$\begin{aligned} R^*Rf(0)&=\int_0^{2\pi} Rf((-\sin\theta,\cos\theta),0)\,d\theta\\ &=\int_0^{2\pi}\int_{-\infty}^{\infty} f(t\cos\theta,t\sin\theta)\,dt\,d\theta\\ &=2\int_{\mathbb{R}^2} f(x)\,|x|^{-1}\,dx. \end{aligned}$$

More generally, the lines through an arbitrary point z are best parametrized $x(t)=z+tw$ (this is a slightly different parametric representation than we gave above where there was no distinguished point on the line, the two differing by an additive change of variable t), so an almost identical computation yields

$$R^*Rf(z)=2\int f(z+x)\,|x|^{-1}\,dx. \tag{2.3}$$

Note that this bears out well what our motivating reasoning led us to expect since the value of f at z is most emphasized in this integral, being multiplied by the singularity of $|x|^{-1}$. This singularity is relatively mild, so there is no difficulty with the convergence of the integral near $x=0$ if f is bounded.

Now the right side of (2.3) is a familiar object. Aside from the constant, it is the convolution of f with the function $|x|^{-1}$. The convolutions with negative powers of $|x|$ are called Riesz transforms, and they are known to be essentially the negative powers of the Laplacian. More precisely, for functions f of $\mathbb{R}^n$, the Riesz transforms are defined

$$I_\alpha f(x)=\gamma_\alpha\int_{\mathbb{R}^n} f(x-y)\,|y|^{\alpha-n}\,dy$$

for $0<\alpha<n$ with

$$\gamma_\alpha=2^{-\alpha}\pi^{-n/2}\frac{\Gamma\left(\frac{n-\alpha}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}.$$

The formal identity $I_\alpha f=(-\Delta)^{-\alpha/2}f$ is best understood via the Fourier transform formula

$$(I_\alpha f)^\wedge(\xi)=|\xi|^{-\alpha}\hat{f}(\xi) \tag{2.4}$$

as compared with

$$(-\Delta f)^\wedge(\xi)=|\xi|^2\hat{f}(\xi) \tag{2.5}$$

where Δ is the Laplacian $\Sigma_{j=1}^n\,\partial^2/\partial x_j^2$. The proof of (2.4), aside from the exact determination of

the constant γ_α, can be deduced from formal properties of the Fourier transform: any convolution operator must multiply the Fourier transforms, the Fourier transform of $|x|^{\alpha-n}$ must be radial and homogeneous of degree $-\alpha$, and therefore a multiple of $|\xi|^{-\alpha}$. For the determination of the constant, and a precise formulation of the above arguments, we refer the reader to [3] or [11].

In terms of the above notation, formula (2.3) says $R^*Rf = 4\pi I_1 f$ on $\mathbb{R}^2$. Referring to (2.4) and (2.5) we can invert I_1 by $I_1(-\Delta)$, so

$$f = \frac{1}{4\pi} I_1(-\Delta) R^* Rf. \tag{2.6}$$

This is our second Radon inversion formula.

3. The Hole Theorem. One aspect of the inversion of the Radon transform that is not apparent from either (2.1), (2.6), or any of the variants of those explicit formulas, is the "hole" theorem. Suppose the function f is defined on the complement of some bounded convex "hole" K. Then Rf is defined for all lines that do not intersect K. Can the values of f outside the hole be recovered from the values of Rf on all lines outside the hole? The hole theorem gives a qualified assent. The qualification is that we must place strong restrictions on the decay of the function at infinity, say that it has compact support. (It is sufficient to assume that f is rapidly decreasing, but the proof we give below does not show this.) In fact the following simple counterexamples show that such restrictions are necessary. Let $f(x_1, x_2) = (x_1 + ix_2)^{-k}$ for any integer k sufficiently large. Then f is defined and continuous in the complement of any neighborhood of the origin, and it can be made to decay at infinity faster than any fixed polynomial rate, although it is not rapidly decreasing. But by the Cauchy integral formula and a simple limiting argument (closing the contour to exclude the origin) the integral of f over any line not passing through the origin is zero. Therefore Rf does not determine f.

In view of the counterexamples, the hole theorem is surprising. In terms of X-rays it can be interpreted as saying that if a convex portion of the object being examined cannot be X-rayed, we can still obtain all the information we need about the rest of the body from X-ray pictures that avoid that portion. (See [9] and [10] for a full discussion of this interpretation.) The object being X-rayed is assumed to be finite, so the hypothesis of compact support is verified.

To prove the hole theorem note that by considering differences it is equivalent to the following

THEOREM. *Let f be a continuous function of compact support. Suppose Rf is zero on every line not intersecting a fixed compact convex set K. Then f is zero outside K.*

We will give a proof that is a simplification of Helgason's proof [5]. (The theorem is sometimes attributed to Ludwig [7], who gave an independent proof, and it also follows from an inversion formula of Cormack [1].) The idea of the proof is to show by induction that any polynomial times f satisfies the same condition. For then we can apply the Weierstrass approximation theorem to every line not intersecting K to conclude that f is zero on that line, hence f is zero outside K. To simplify notation let us assume the line is vertical, $x_1 = s$ fixed, $s > 0$ and that K lies to the left of the line. Again for simplicity we will show that for any polynomial p of degree one, the integral of pf along the line is zero, for then this argument can be used for the induction step.

Actually it suffices to handle the polynomial $p(x_1, x_2) = x_2$, since $a + bx_1$ is constant along the line $x_1 = s$. In other words

$$\int_{-\infty}^{\infty} tf(s,t)\,dt = 0. \tag{3.1}$$

The gist of the proof is that (3.1) can be established by "rocking the boat." We look at lines tangent to the circle of radius s about the origin. These lines are parametrized by

$$x(t) = (s\cos\theta + t\sin\theta, -s\sin\theta + t\cos\theta).$$

For $\theta = 0$ this is the vertical line $x_1 = s$. Because K is compact and convex the lines $x(t)$ do not

intersect K for θ sufficiently close to zero. Thus we know that the integral of f over these lines is equal to zero:

$$\int_{-\infty}^{\infty} f(s\cos\theta + t\sin\theta, -s\sin\theta + t\cos\theta)\,dt = 0.$$

Now we "rock the boat" by differentiating this identity with respect to θ and then setting $\theta = 0$. The result is

$$\int_{-\infty}^{\infty}\left(t\frac{\partial f}{\partial s}(s,t) - s\frac{\partial f}{\partial t}(s,t)\right)dt = 0.$$

However $\int_{-\infty}^{\infty}(\partial f)/(\partial t)(s,t)\,dt = 0$ by the fundamental theorem of the calculus, since f has compact support, so

$$\int_{-\infty}^{\infty} t\frac{\partial f}{\partial s}(s,t)\,dt = 0. \tag{3.2}$$

At first glance it appears that we have failed to obtain (3.1) because of the s-derivative in (3.2). However there is a magic trick for making the derivative disappear. Recall that the convex set K lies to the left of the line $x_1 = s$. This means that the same argument can be applied to any line $x_1 = r$ for $r \geqslant s$. Integrating (3.2) over all such lines we obtain

$$\int_{s}^{\infty}\int_{-\infty}^{\infty} t\frac{\partial f}{\partial r}(r,t)\,dt\,dr = 0.$$

Finally we interchange the order of integration, and apply the fundamental theorem of calculus to obtain (3.1), using the compact support of f to justify both steps.

The above argument used the differentiability of f, but a routine regularization and limiting argument allows us to dispense with this extra assumption. Also the argument generalizes easily to higher dimensions.

The hole theorem can be thought of as a Radon inversion theorem for functions defined on the complement of the hole. It is only a conditional inversion because of the support assumption.

This is analogous to the fact that a function of compact support on the line can be recovered from its derivative by the fundamental theorem $f(x) = \int_{-\infty}^{x} f'(t)\,dt$, but for f constant this is no longer valid. In some of the problems below, the inversion formulas will also be conditional; in others they will be valid whenever they make sense.

4. Variations. We have seen above that the strategy of looking at R^*R leads to an explicit inversion of the Radon transform. There are many other situations where this strategy is useful, and it is instructive to look at a whole class of examples. Consider a fixed "shape" S_0 in the plane, and then all congruent shapes, the images of S_0 under rigid motions. Assuming that there is some natural way of integrating functions over S_0 (technically, we want a measure supported on S_0, but in the examples it will always be obvious what to do), the same integration can be performed over the congruent images. If we let $\mathbb{S}$ denote the class of congruent images, and S an arbitrary element of $\mathbb{S}$, then

$$Rf(S) = \int_S f$$

is the associated Radon transform. If $\mathbb{S}$ is the class of straight lines, then we have the ordinary Radon transform. But we could choose for $\mathbb{S}$ the class of circles of fixed radius, squares of fixed size (or their boundaries), rectangles, triangles, etc. There is no need to consider only the plane, since the notion of congruence and rigid motion makes sense in higher dimensions. But for simplicity let us restrict attention to the plane and allow only orientation preserving rigid motions. These are of the form

$$x \to a(\theta)(x+b) \quad \text{where } a(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

is a rotation through angle θ, and b is an arbitrary plane vector. We can parametrize the set $\mathbb{S}$ of congruent images of S_0 by $a(\theta)$ and b, by setting $S(a(\theta), b) = \{x : a(\theta)(x+b) \in S_0\}$. This is not a canonical labeling since there is no distinguished S_0, and it may not be one-to-one if S_0 has any symmetries; but no matter. We write

$$\begin{aligned} Rf(a(\theta), b) &= \int_{S(a(\theta),\, b)} f \\ &= \int_{S_0} f(a(-\theta)y - b)\, d\mu(y) \\ &= \int_{S_0} f(y-b)\, d\mu(a(\theta)y) \end{aligned}$$

where we let $d\mu(y)$ stand for the measure of integration on S_0. Writing $d\mu_\theta(y) = d\mu(a(\theta)y)$ for the rotated measure, we can interpret this expression in a significant way: *for each fixed* θ, $Rf(a(\theta), b)$ *is the convolution of* $f(-x)$ *with* $d\mu_\theta$ *at the point* b.

The problem of when this Radon transform is invertible is known as the Pompeiu problem (see Zalcman [**12**] and the references there). If no growth conditions are put on the function f, then the problem becomes quite technical. Therefore we will assume that f is bounded by a polynomial. With this assumption, the Fourier transform is the natural tool to illuminate the problem. Indeed since the Fourier transform of a convolution is the product of Fourier transforms, the Fourier transform in the b-variable of $Rf(a(\theta), b)$ is

$$\begin{aligned} \int_{\mathbb{R}^2} Rf(a(\theta), b) e^{-ib\cdot\xi}\, db &= \hat{f}(-\xi)\hat{d\mu}_\theta(\xi) \\ &= \hat{f}(-\xi)\, \hat{d\mu}(a(-\theta)\xi). \end{aligned}$$

Here we have used the fact that the Fourier transform of the rotation of $d\mu$ is the rotation (through the negative angle) of the Fourier transform of $d\mu$. For fixed ξ the points of $a(-\theta)\xi$ describe a circle about the origin, so if $\hat{d\mu}(\xi)$ does not vanish identically on any circle about the origin, we can recover $\hat{f}$ from Rf and hence f by the Fourier inversion formula. (It never vanishes at the origin since $\hat{d\mu}(0) = \int_{S_0} d\mu > 0$.) Normally we would not expect a random function on the plane to vanish on a circle unless there is some good reason for it to do so— for instance if it is radial. An example of this is produced by taking S_0 to be the unit circle. Then $\hat{d\mu}$ is known exactly (see [**3**] or [**11**]),

$$\begin{aligned} \hat{d\mu}(\xi) &= \int_0^{2\pi} e^{-i(\xi_1\cos\theta + \xi_2\sin\theta)}\, d\theta \\ &= J_0(|\xi|) \end{aligned}$$

where J_0 is the Bessel function of order zero. Since it is known that the Bessel function has isolated zeroes, the Radon transform is not invertible. (In fact, $J_0(\lambda^{-1}|x|)$, where λ is any such zero, is a nonzero function whose integrals over all circles of radius one are zero.) Of course in this case we can still hope for a conditional inversion since Rf determines $\hat{f}$ except on isolated circles, so if we assume say $f \in L^1$ or L^2, f is determined. (Of course neither condition is at all natural for the problem.) Such a conditional inversion problem is discussed in John [**6**], but we shall not consider it further here. Let us observe, however, that no such problem arises if we let $\mathbb{S}$ be the class of *circles* of radius one in three-space. The Fourier transform of the circle measure is constant in the direction perpendicular to the circle, hence it vanishes on isolated cylinders, and it is an obvious geometric fact that no sphere is contained in a countable union of cylinders!

Next let's look at what the dual transform does for us in this context. Obviously we want to define R^* by integrating over all sets $S \in \mathbb{S}$ which contain a fixed point x. If we fix the angular variable θ, then $x \in S(a(\theta), b)$ if and only if $b \in S(a(\theta), x)$; in other words the set of all values

of the b parameters for which $S(a(\theta), b)$ contains x is the set $S(a(\theta), x)$, so it is natural to define

$$R^*g(x) = \int_0^{2\pi}\left(\int_{S(a(\theta),\,x)} g(a(\theta), b)\right) d\theta$$
$$= \int_0^{2\pi}\int_{S_0} g(a(\theta), y - x)\, d\mu(a(\theta)y)\, d\theta.$$

Then we compute

$$R^*Rf(x) = \int_0^{2\pi}\int_{S_0} Rf(a(\theta), y - x)\, d\mu(a(\theta)y)\, d\theta$$
$$= \int_0^{2\pi}\int_{S_0}\int_{S_0} f(z + x - y)\, d\mu(a(\theta)z)\, d\mu(a(\theta)y)\, d\theta.$$

If we ignore the θ integration, the expression we have is the triple convolution $f * d\mu_\theta * \widehat{d\mu_\theta(-x)}$ which goes under Fourier transformation into the triple product $\hat{f}(\xi)\, \hat{d\mu}_\theta(\xi)\, \widehat{d\mu_\theta(-\xi)}$. But since $d\mu_\theta$ is real-valued $\widehat{d\mu_\theta(-\xi)} = \hat{d\mu}_\theta(\xi)$, so

$$\int R^*Rf(x) e^{-ix\cdot\xi}\, dx = \int_0^{2\pi} \hat{f}(\xi)\, |\, \hat{d\mu}(a(-\theta)\xi)\, |^2\, \hat{d}\theta$$
$$= \hat{f}(\xi)\int_0^{2\pi} |\, d\mu(a(\theta)\xi)\, |^2\, d\theta.$$

Now observe that $m(\xi) = \int_0^{2\pi} |\, \hat{d\mu}(a(\theta)\xi)\, |^2\, d\theta$ is nonvanishing if and only if $\hat{d\mu}(\xi)$ does not vanish identically on any circle about the origin. Thus we lose no information in considering R^*Rf instead of Rf; if we can express f in terms of Rf, then we can express it in terms of R^*Rf, and the expression will be simpler. In fact, by the Fourier inversion formula

$$f(x) = \frac{1}{(2\pi)^2}\int m(\xi)^{-1}\int R^*Rf(y) e^{-iy\cdot\xi}\, dy e^{ix\cdot\xi}\, d\xi$$

and $m(\xi)$ is a radial function. In some cases it may be possible to further simplify this expression, but in the examples in problem **E** I have not seen any way to do this.

5. Problems.

A. Consider any *finite plane geometry*, that is, a finite set X of points and a set Y of lines, each line $y \in Y$ being a subset of X subject to the single axiom "two points determine a unique line." In other words, for any two points $x_1, x_2 \in X$, there exists a unique $y \in Y$ such that $x_1 \in y$ and $x_2 \in y$. For a function f on X, define the Radon transform $Rf(y) = \Sigma_{x\in y} f(x)$. Can you find an explicit inversion formula for f in terms of Rf? Assume that there are at least two lines.

There is a natural generalization of this problem to *finite n-dimensional geometries*. Such a geometry is a finite set X_0 of points, and sets X_k of k-planes, $k = 1,\dots, n-1$, such that each k-plane $x_k \in X_k$ is a subset of X_0. We need two axioms: (1) any $k+1$ points belong to a unique k-plane; (2) if $k+1$ points in a k-plane x_k belong to a single m-plane x_m for $m > k$, then all points of x_k belong to x_m. The k-plane transform for fixed k is then defined as $R_k f(x_k) = \Sigma_{x\in x_k} f(x)$. Can you invert this transform? Assume that there are at least two k-planes for each k.

B. Let L denote the lattice of integer points in the plane, points (k, m) where k and m are integers. Let f be any absolutely summable function on L; in other words

$$\sum_{k=-\infty}^{\infty}\ \sum_{m=-\infty}^{\infty} |f(k, m)|$$

is finite. Let Y denote the set of lines in the plane which intersect the lattice in more than one

point. Then it makes sense to define the Radon transform

$$Rf(y) = \sum_{(k,m)\in y} f(k,m),$$

the series being absolutely convergent. Can you invert this transform?

Whatever solution you find, you will have to use the hypothesis that f is absolutely summable, because there exist nonzero functions f such that the series for Rf is absolutely convergent for each line and $Rf \equiv 0$. Can you find such a function?

C. Let N denote the positive integers. For an absolutely summable function on N, define a Radon transform

$$Rf(m) = \sum_{k=1}^{\infty} f(km)$$

for all $m \in N$. Can you invert this transform, assuming some more rapid decay of f, say $|f(n)| \leqslant cn^{-2-\varepsilon}$?

D. Let T^2 denote the torus, which is the plane modulo the lattice L of problem **B**. Some lines in the plane, when projected on the torus, wrap around infinitely often, and others come back upon themselves after a finite length. These can be distinguished according as the slope is irrational or rational. Using the lines of finite length, we can define the Radon transform of a continuous function f on T^2 to be the mean value of f on each such line. (It is actually possible to define the mean value on the lines of infinite length using the Bohr mean from the theory of almost periodic functions, but the mean will always be the usual mean value of f on T^2, and this information is available from the mean value of f on say all horizontal lines.) Can you invert this transform? Part of the problem is to choose appropriate representations for the finite lines on T^2. You may assume that f has an absolutely convergent double Fourier series.

E. Let $\mathbb{S}$ be the set of squares (interiors) with side length one in the plane, and $Rf(S) = \int\int_S f(x_1, x_2)\, dx_1\, dx_2$. Can you prove that R is invertible? Suppose instead of the interiors of the squares we consider the boundaries of the squares, with $Rf(S)$ equal to the line integral? Finally, what if we consider the vertices of the square?

References

1. A. M. Cormack, Representation of a function by its line integrals, with some radiological applications I, J. Appl. Phys., 34 (1963) 2722–2727.

2. P. Funk, Über eine geometrische Anwendung der Abelschen Integralgliechung, Math. Ann., 77 (1916) 129–135.

3. I. M. Gelfand and G. E. Shilov, Generalized Functions, vol. 1, Academic Press, 1964.

4. C. Hamaker, K. T. Smith, D. C. Solmon, and S. L. Wagner, The divergent beam X-ray transform, Rocky Mountain J. Math., 10 (1980) 253–283.

5. S. Helgason, The Radon Transform, Birkhauser, 1980.

6. F. John, Plane waves and spherical means applied to partial differential equations, Interscience, 1955.

7. D. Ludwig, The Radon Transform on Euclidean space, Comm. Pure Appl. Math., 23 (1966) 49–81.

8. J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, Ber. Verh. Sachs. Akad. Wiss. Leipzig, Math-Nat. Kl., 69 (1917) 262–277.

9. L. A. Shepp and J. B. Kruskal, Computerized tomography: the new medical X-ray technology, this MONTHLY, 85 (1978) 420–439.

10. K. T. Smith, D. C. Solmon, and S. L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, Bull. Amer. Math. Soc., 83 (1977) 1227–1270.

11. E. M. Stein and G. Weiss, Introduction to Fourier analysis on Euclidean spaces, Princeton University Press, 1971.

12. L. Zalcman, Offbeat Integral Geometry, this MONTHLY, 87 (1980) 161–175.

The solutions of the problems occur on page 420.

MEANING AND INFORMATION IN CONSTRUCTIVE MATHEMATICS

FRED RICHMAN
Department of Mathematics, New Mexico State University, Las Cruces, NM 88003

The publication of Errett Bishop's *Foundations of Constructive Analysis* presented the working mathematician with a viable alternative to the prevailing interpretation of mathematical statements. The interpretation suggested by Bishop is not wholly new; many of the ideas go back to L. E. J. Brouwer in the early part of this century, and have been refined and embellished by the Dutch intuitionists. But these ideas were never presented in a manner that was convincing to the average mathematician. Their exposition was often accompanied by an implication that classical mathematics was at best built on insecure foundations, if not outright false. The call was for salvage operations, and the recommended action included forbidding the use of widely accepted arguments. People who felt perfectly comfortable with the mathematics they were doing were unimpressed by these appeals. They saw no reason to sacrifice their powerful mathematical tools and theories on the altar of some crackpot notion of legitimacy.

Bishop showed that one could adopt a thoroughgoing constructive point of view and still do mathematics as it is usually understood. He did this by appropriating standard mathematical symbolism to carry constructive meaning, rather than introducing a specialized notation, and by developing large areas of rather sophisticated mathematics in a constructive manner. His book can be appreciated by mathematicians unfamiliar with logic or recursive function theory, and avoids the more bizarre intuitionistic notions of choice sequence and bar induction.

But Bishop-style constructivists also tend to be polemical, and they are met with some of the same hostility and indifference that Brouwer was. Possibly this situation will improve as the constructivists master their trade and as the revolution in mathematical consciousness, brought about by the advent of the high-speed digital computer, spreads. In the meantime it seems appropriate to try to convey to mathematicians, and to those interested in the philosophy of mathematics, some understanding of what constructive mathematics means to its practitioners. In what follows I shall attempt to do this via a case study. Our starting point is the following seemingly trivial theorem:

THEOREM A. *Let $\mathbb{R}$ denote the set of real numbers and let $x \in \mathbb{R}$. Then the set $\mathbb{R}x$ of multiples of x is closed.*

Proof. If $x = 0$, then $\mathbb{R}x$ is the set whose only member is 0. This set is closed—the limit of a sequence of 0's is 0. On the other hand, if $x \neq 0$, then $\mathbb{R}x = \mathbb{R}$, which is closed since it contains every possible limit point. □

The proof is somewhat inelegant because of the division into cases: one argument when $x = 0$, one when $x \neq 0$. The reader is invited to attempt a proof that avoids this artifice. The cases are dramatically different; the set $\mathbb{R}x$ changes suddenly from a single point to the whole space when x moves away from 0. But this only says that the *argument* is discontinuous as a function of the data, not that this discontinuity is an essential feature of the theorem. The theorem that $\mathbb{R}x$ is a *vector space* can also be proved by considering the two cases $x = 0$ and $x \neq 0$, but this would clearly be the wrong way to proceed. So we might still hope to discover a proof that makes no reference to whether x is 0 or not.

How would such a proof go? We are given a real number y and a sequence of real numbers r_n such that $r_n x$ converges to y. We are required to find a real number r_∞ such that $y = r_\infty x$. If x is 0

Fred Richman received his Ph.D. from the University of Chicago in 1963 under the direction of Irving Kaplansky. He has written a book on elementary number theory and coauthored texts on trigonometry and mathematics for the liberal arts student. Besides constructive mathematics his main research area is infinite abelian group theory. A Life Master in bridge, he plays a little piano and less violin.

we can choose r_∞ arbitrarily; if $x \neq 0$ we let r_∞ be y/x. The problem is to find r_∞ in a manner that is independent of whether $x = 0$.

Why would anyone care to find such a proof? We have mentioned the lack of elegance of a proof by cases, and it can be argued that elegance is a companion of comprehensibility and insight. Were we to find a proof that was indifferent to whether or not $x = 0$, then we might claim that our previous proof did not reveal the real reason why the theorem was true. I think that most mathematicians would agree that an unnecessary split of a proof into cases is, more often than not, a barrier to understanding.

Another reason for attempting to eliminate the two cases in this proof concerns the possibility of our not being able to determine whether x is 0 or not. If we assume the standard omniscient attitude toward mathematical objects, then this possibility makes no sense. However if we imagine that not all information about the number x is available to us, then we can entertain such a possibility. We postpone the question as to why we might want to put ourselves in such a frame of mind. For now let's try to pin down just what information should be available to us, and what sort need not be.

Imagine that our link with the universe of mathematical objects consists of a computer terminal. Dealing with rational numbers will be relatively unproblematic as they are specified by pairs of integers, each integer being represented by a string of digits. Real numbers are a different story: To specify a real number requires an infinite sequence of rational numbers. One possibility is to represent sequences by programs of functions f such that $f(n)$ is the nth term of the sequence in question. However the identification of real numbers with computer programs does violence to the conventional concept of a real number; for example, computer programs are countable, but real numbers are not (Richard's paradox). In all fairness it must be pointed out that there are plausible arguments in favor of this identification, and that the Russian constructivists have adopted this point of view.

I prefer to make the minimum commitment as to the nature of real numbers. By employing an operational definition we can accommodate the widest range of viewpoints. Whatever we mean by a real number we must be able, for each positive integer n, to produce a rational number a_n that approximates that real number to within $1/n$. These numbers a_n may in fact be generated by some computer program, but that is irrelevant. What counts is that any one of them is available to us on demand. We may think of them as being produced by a library routine based on an algorithm that we do not necessarily know.

If we are given a real number x in this fashion, then it is clear that we may not have enough information to decide whether x is 0 or not. If we find out that $a_{1000} = 0.01$, then we know that $x > 0$. But if $a_{1000} = 0$, then we can neither say that $x = 0$ nor say that $x \neq 0$. And we will be in the same boat no matter how many a_n we find out are 0.

Having indicated why we may be loath to use the fact that $x = 0$ or $x \neq 0$ in the proof of Theorem A, we can see that we are dealing with something more than a question of taste in proofs—it is a question of meaning. From the constructive point of view Theorem A was misstated; it should have read:

THEOREM B. *If $x = 0$ or $x \neq 0$, then $\mathbb{R}x$ is closed.*

This more modest theorem is what was proved. The reason the extra hypothesis was left off is that, in the conventional interpretation, the statement "$x = 0$ or $x \neq 0$" is a tautology. In the constructive interpretation it entails that we can determine which of the two alternatives holds.

We might pause here to observe the difference in information content between the statements "$x = 0$" and "$x \neq 0$". To establish that a real number x given by a sequence of rational approximations a_n is different from 0, we must find n such that $|a_n| > 1/n$. To establish that $x = 0$, we must show that $|a_n| \leqslant 1/n$ for each n, or, equivalently, that x cannot be different from zero. Thus the statement "$x \neq 0$" is positive in that it requires the specification of n and the corresponding rational number $|a_n| - 1/n$ that bounds x away from 0. The statement "$x = 0$" is

essentially negative: it is equivalent to the denial of "$x \neq 0$". On the other hand, knowing "$x = 0$" is false does not provide the data necessary to establish "$x \neq 0$" (although the Russian constructivists argue that this data is available).

Once we have rephrased Theorem A in the form of Theorem B, the proof-theoretic question of whether we can avoid a division into cases becomes a question about truth: Is Theorem A true? And we have the possibility of a definitive negative answer, from the constructive point of view, by proving the converse of Theorem B:

THEOREM C. *If* $\mathbb{R}x$ *is closed, then* $x = 0$ *or* $x \neq 0$.

To establish Theorem C we must show how, from the information that $\mathbb{R}x$ is closed, we can come to a decision regarding whether x is 0 or not.

What sort of information is involved in knowing that $\mathbb{R}x$ is closed? A subset S of R is closed if, given any y in $\mathbb{R}$, together with the ability to construct for each n a number s_n in S such that s_n is within $1/n$ of y, we can establish that $y \in S$. For $S = \mathbb{R}x$, the latter means that we must be able to construct r in $\mathbb{R}$ such that $y = rx$. So our knowledge that $\mathbb{R}x$ is closed entails our possession of a procedure for constructing r from the numbers s_n and y. Specification of y is in fact superfluous as y is the unique limit of the sequence s_n and can be constructed given any sequence s_n such that

$$|s_n - s_m| \leqslant 1/n + 1/m$$

for all n and m, which clearly holds for the given sequence.

Thus the problem is of a familiar, even prototypical, variety. We have a "machine" that converts certain kinds of data (Cauchy sequences s_n in $\mathbb{R}x$) into another kind (real numbers r). We are interested in a third kind of information (is $x = 0$ or is $x \neq 0$?). We must construct a clever Cauchy sequence s_n so that the resulting r will enable us to resolve the question "is $x \neq 0$?".

Although we cannot always determine whether a given number x is different from zero, we can show that x is either small, or bounded away from zero. More generally, if $a < b$ are rational numbers, then we can show that either $x < b$ or $x > a$. Indeed let x' be a rational number within $(b - a)/3$ of x. If $x' \leqslant (b + a)/2$, then $x < b$; while if $x' \geqslant (b + a)/2$, then $x > a$. With this in mind we can prove Theorem C.

Proof of Theorem C. Choose $\lambda_1, \lambda_2, \lambda_3, \ldots$ such that

(i) each λ_n is either 0 or 1,

(ii) if $\lambda_n = 1$, then $|x| < 1/n^2$,

(iii) if $\lambda_n = 0$, then $|x| > 1/(n + 1)^2$.

Set $r_n + \Sigma_{i=1}^{n} \lambda_i$. Then $r_n x$ is a Cauchy sequence of real numbers as

$$(r_n - r_{n-1})x = \lambda_n x = \begin{cases} 0 & \text{if } \lambda_n = 0 \\ x & \text{if } \lambda_n = 1 \end{cases}$$

has absolute value less than $1/n^2$. So $r_n x$ converges to a real number, which is in $\mathbb{R}x$ since $\mathbb{R}x$ is closed. Thus we can write its limit as $r_\infty x$ for some real number r_∞. Let N be a positive integer exceeding r_∞. If $\lambda_N = 0$, then $|x| > 1/(N + 1)^2$ so $x \neq 0$. If $\lambda_N = 1$, then $r_N = N$; so if $|x| > 0$, then $r_\infty \geqslant N$, a contradiction. Thus $x = 0$. □

What is the point of Theorem C? It seems rather doubtful that one would ever be in a position to verify the hypothesis without already having established the conclusion. But there are many theorems in mathematics of this kind, and not all of them are bad. By proving the converse of a theorem one shows that the hypothesis was necessary, that we cannot get a more general theorem by a genuine weakening of the hypothesis. Thus such a theorem serves the same function as a counterexample in delimiting the range of validity of possibly useful theorems.

The reader who understands the content of Theorem C understands the constructive approach

to mathematics. There remains the question as to why anyone should choose this approach. We could investigate the claims that it blends an applied mathematics mentality with the subject matter of pure mathematics, and that this is good; or that it more accurately reflects the *doing* of mathematics; or that it directs mathematical research to more significant areas. In the last analysis, however, these arguments will shed little light on the issue. There it is—if it feels good, do it.

The following incomplete bibliography is provided for those who wish to acquaint themselves with recent work in constructive mathematics. A rather extensive bibliography can be found in Bridges' book [3].

References

1. G. Berg, W. Julian, R. Mines and F. Richman, The constructive equivalence of covering and inductive dimension, Gen. Top. and its Applications, 7(1977) 99–108.
2. E. Bishop, Foundations of Constructive Analysis, McGraw-Hill, New York, 1967.
3. D. S. Bridges, Constructive Functional Analysis, Research Notes in Mathematics, 28, Pitman, London, 1979.
4. Y. K. Chan, "On some open problems in constructive probability theory," in F. Richman, editor, Constructive Mathematics, Proceedings of the 1980 Las Cruces conference, Lecture Notes in Mathematics 873, Springer-Verlag, Berlin, 1981.
5. H. Chen, A constructive Riemann mapping theorem, Pacific J. Math., 44(1973) 435–454.
6. W. Julian, R. Mines and F. Richman, Algebraic numbers, a constructive development, Pacific J. Math., 74(1978) 91–102.
7. M. Mandelkern, Located sets on the line, Pacific J. Math., vol. 93, no. 1, 1980.
8. J. Nuber, A constructive ergodic theorem, Trans. Amer. Math. Soc., 164(1972) 115–137.
9. F. Richman, The constructive theory of countable abelian p-groups, Pacific J. Math., 45(1973) 621–637.
10. L. Rogers, Basic subgroups from a constructive viewpoint, Communications in Algebra, 8(1980) 1903–1925.
11. A. Seidenberg, Constructions in algebra, Trans. Amer. Math. Soc., 197(1974) 273–313.

MISCELLANEA

78. It is the most difficult and the most important task of the mathematical teacher that he should strive to attain a broad view of his subject, and to discern the main roads of its development right down to the new work that is being carried on to-day. Only so can he hope to lead his students all the way along those roads, and only so can he worthily initiate them into the mysteries of our craft. An honours course which stops where mathematics was a century ago is unworthy of a living subject. It may provide some useful material for the study of the sciences, though even this is becoming more doubtfully true. It will not give the student any adequate appreciation of mathematics as one of the greatest of the achievements of the human intellect.

By an inflexible determination not to be drawn aside into the attractive side roads of over-generalization or trick problem solving, the teacher may hope to lead his students all the way along some of these main roads, but with the great scope of our subject and with the limitations of time he will be compelled to choose even between the main roads. I have no hesitation in proposing as a principle of selection that he should choose those main roads which lead right into the work of the greatest living mathematicians. Mathematics is not a closed body of theory. It is the creative activity of men and women. It is the work of those who are filled with an insatiable curiosity to know the answers to its unsolved problems, who know what it is to wrestle with these problems and to fail, who sometimes know the joy of finding at least a partial answer to some of them. This is the living spirit of mathematics which we are proud to find so vital in our own Society. Unless the student can be brought into some direct contact with this living thing, he has lost all that is of most value in what we have to give him. When the teacher has to choose between different important branches of mathematics, he will therefore do well to give greater prominence to those branches which are the necessary preliminaries to the study of the best present day work.

—G. B. Jeffery, Presidential Address (1937) to the London Mathematical Society (*J. London Math. Soc.*, 13 (1938), p. 73).

CENTER SECTION

(Vol. 89, No. 6, June-July 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General. A Dictionary of Statistical, Scientific and Technical Terms. Hardeo Sahai, José Berríos. Wadsworth, 1981, ix + 143 pp, $13.95. [ISBN: 84-534-0004-0] English-Spanish and Spanish-English dictionary of terms from statistics and related fields. Each part contains about 3000 entries including phrases (e.g., joint frequency distribution) and terms with proper names (e.g., Mann-Whitney U-test). KS

General, S. Puzzle It Out: Cubes, Groups and Puzzles. John Ewing, Czes Kośniowski. U of Cambridge Pr, 1982, 64 pp, $4.95 (P). [ISBN: 0-521-28924-6] An elementary full-color exposition of the group theory of the cube and related puzzles (15 puzzles, cylinders, tetrahedrons), illustrating and applying cycles, conjugations and commutators. Includes a new two-dimensional figure-of-eight puzzle. A marvelous supplement to any introductory course on group theory, although the writing is geared more to a high school than a college audience. LAS

Education, P, L. Learning with Computers. Alfred Bork. Digital Pr, 1981, xvi + 286 pp, $25. [ISBN: 0-932376-11-8] Reprints of papers published in diverse sources during the past decade, reporting on projects of the Physics Computer Development Project and the Educational Technology Center at the University of California at Irvine. The eclectic sources impose certain handicaps: several articles are clearly dated; others are like local user guides. Nevertheless, the volume as a whole reflects imaginative use of computers in education, and is well worth examining. LAS

Foundations, P. Lecture Notes in Mathematics-834: Model Theory of Algebra and Arithmetic. Ed: L. Pacholski, J. Wierzejewski, A.J. Wilkie. Springer-Verlag, 1980, vi + 410 pp, $24.50 (P). [ISBN: 0-387-10269-8] Proceedings of the Conference on Applications of Logic to Algebra and Arithmetic held at Karpacz, Poland, September 1-7, 1979 plus some papers by mathematicians who were invited but could not attend. KS

Complex Analysis, T*(18), S*, P*. Function Theory of Several Complex Variables. Steven G. Krantz. Wiley, 1982, xiii + 437 pp, $39.95. [ISBN: 0-471-09324-6] This useful and engagingly written text "introduces the classically-oriented analyst to holomorphic functions on C^n," and contains a wealth of recent results and typical computations never before collected in book form. The bias is function-theoretic: sheaf theory, e.g., is treated only briefly. Attractive features: many exercises, open problems, analogies and failures of analogy between the one- and several-variable settings. PZ

Differential Equations, T(14), S, L. Ordinary Differential Equations and Stability Theory: An Introduction. David A. Sánchez. Dover, 1979, viii + 164 pp, $3 (P). [ISBN: 0-486-63828-6] An introduction to ordinary differential equations that takes the point of view that it is not necessary to learn many specialized techniques for solutions of special equations before one can understand some of the qualitative features of differential equation theory. Discusses the concept of stability as it relates to autonomous and non-autonomous equations. W.H. Freeman 1968 edition (TR, January 1969). JG

Numerical Analysis, S(17), P. Multi-Dimensional Continued Fraction Algorithms. A.J. Brentjes. Math. Centre Tracts, V. 145. Math Centrum, 1981, 183 pp, Dfl. 23,10 (P). [ISBN: 90-6196-231-5] Continued fraction algorithms can be used to find all best approximations to a line in n-space. Application to Diophantine approximation problems. JG

Functional Analysis, P. Dimensions and C*-Algebras. Edward G. Effros. CBMS Reg. Conf. in Math., No. 46. AMS, 1981, v + 74 pp, $7.20 (P). [ISBN: 0-8218-1697-7] Monograph based on lectures presented at CBMS regional conference held at Oakland University, Rochester, Michigan, June 25-29, 1979. Provides introduction to C*-algebraic K_0 theory; primarily concerned with dimensions in AF (approximately finite) algebras. KS

Analysis, T(18), S, P. Random Differential Inequalities. G.S. Ladde, V. Lakshmikantham. Math. in Sci. and Eng., V. 150. Academic Pr, 1980, xi + 211 pp, $30. [ISBN: 0-12-432750-8] A systematic treatment of random differential inequalities through each of sample calculus, LP-mean calculus, and Itô-Doob calculus. Lack of few exercises limit its use as a text, but would be a good reference. PH

Analysis, P. Lecture Notes in Mathematics-886: Fixed Point Theory. Eds: E. Fadell, G. Fournier. Springer-Verlag, 1981, xii + 511 pp, $27 (P). [ISBN: 0-387-11152-2] Proceedings of a June 1980 conference held at Sherbrooke, Quebec. LAS

Algebraic Geometry, P. Blowing Up Grassmannians. Ari Babakhanian. Pure and Appl. Math., No. 59. Queen's U, 1981, 77 pp, (P). A blow-up is a monoidal transformation. JG

Algebraic Geometry, P. Abel-Jacobi Isogenies for Certain Types of Fano Threefolds. G.E. Welters. Math. Centre Tracts, V. 141. Math Centrum, 1981, 139 pp, Dfl. 17,85 (P). [ISBN: 90-6196-227-7] The author's doctoral dissertation; the main results being that for certain varieties in P^3 or P^3 the so-called Abel-Jacobi map is either an isomorphism or an isogeny. SG

Geometry, S(15-16), P. Lecons de Géométrie: Ier semestre: Géométrie Analytique. Mikhaiil Postnikov. MIR, 1981, 279 pp. Linear algebra leads to classical analytic geometry and then to affine and more general geometric considerations. A sophisticated presentation with no exercises, containing lots of things teachers, at least, should read about. JAS

Geometry, S, P. Curves and Symmetry, Volume 1. J. Lee Kavanau. Science Software Systems, 1982, xviii + 430 pp, $19.95 (P). [ISBN: 0-937292-01-X] This typescript edition supplies the illustrations and amplifies the material for Chapters 6-9 in the author's 1980 Symmetry, An Analytic Treatment (TR, February 1981). JNC

Algebraic Topology, P. Continuous Cohomology of the Lie Algebra of Vector Fields. Toru Tsujishita. Memoirs No. 253. AMS, 1981, iv + 154 pp, $9.20 (P).

Statistics, S(13-17). Statistical Tables, Second Edition. F. James Rohlf, Robert R. Sokal. W H Freeman, 1981, xiii + 219 pp, $22; $9.95 (P). [ISBN: 0-7167-1257-1; 0-7167-1258-X] Revision of the authors' 1969 book of tables (TR, October 1969 and May 1970). Contains 13 new tables, including 5 for multiple comparisons, 2 for Kolmogorov-Smirnov statistics and 2 for testing outliers, while eliminating tables of simple mathematical functions. Explanations and applications of the tables can be found in Sokal and Rohlf's Biometry, Second Edition (TR, May 1982). RSK

Statistics, T(16-18: 1), S, L. Time Series Models. A.C. Harvey. Halsted Pr, 1981, x + 229 pp, $39.95. [ISBN: 0-470-27259-7] Stationary stochastic processes, the time domain, the frequency domain, state space models, the Ralman filter, estimation for autoregressive-moving average models, prediction. Stresses applications rather than proofs. Presupposes knowledge of statistics, calculus, and linear algebra. FLW

Computer Literacy, T(13: 1). Computers and Business Information Processing. William S. Davis. Addison-Wesley, 1981, xviii + 475 pp. [ISBN: 0-201-03161-2] Thorough and interesting introduction to computer literacy concepts from a business perspective. Student workbooks, instructor's manual, test item bank, and transparency masters are available. JJ

Computer Programming, S, P. Some Common BASIC Programs: Atari Edition. Lon Poole, Mary Borchers, Steven Cook. Osborne/McGraw-Hill, 1981, 200 pp, $14.99 (P) [ISBN: 0-931988-53-5]; Some Common BASIC Programs: Apple II Edition. Lon Poole, Mary Borchers, David M. Castlewitz. [ISBN: 0-931988-68-3] Collection of source listings of 76 mathematically-oriented programs written in ATARI and Applesoft Basic. Several involve the use of a printer. Areas of emphasis are business mathematics, precalculus, calculus, and statistical tests. JJ

Computer Programming, S, P. Explore Computing with the TRS-80 (& Common Sense) with Programming in BASIC. Richard V. and Josephine P. Andree. Prentice-Hall, 1982, x + 230 pp, $11.95 (P). [ISBN: 0-13-296137-7] Classroom-tested materials for the active development of programming skills at the secondary school level, especially using TRS-80 Basic. Best feature of the text is its extensive collection of "micro-research" problems, most of them with a mathematical bent. JJ

Computer Science, P, L. A User Guide to the UNIX System. Rebecca Thomas, Jean Yates. Osborne/McGraw-Hill, 1982, xi + 508 pp, $15.99 (P). [ISBN: 0-931988-71-3] The first commercial guide to Bell Labs UNIX operating system, now used by 90% of university computer science departments. Includes tutorials on common commands, general comments on using UNIX in office automation, and lists of UNIX resources (hardware, software, related systems, user groups). Provides a good overview, but lacks sufficient detail to serve as a complete reference manual. LAS

Systems Theory, P. Stochastic Systems: The Mathematics of Filtering and Identification and Applications. Eds: Michiel Hazewinkel, Jan C. Willems. D. Reidel Pub, 1981, xi + 663 pp, $69.50. [ISBN: 90-277-1330-8] Proceedings of the June 1980 NATO Advanced Study Institute held at Les Arcs, Savoire, France, enriched with tutorials to make the volume attractive to a diverse audience (e.g., statisticians, geologists, econometricians, mathematicians). A complete survey from basic graduate school techniques to research frontiers. LAS

Applications (Artificial Intelligence), P. From Images to Surfaces: A Computational Study of the Human Early Visual System. William Eric Leifur Grimson. MIT Pr, 1981, xiii + 274 pp, $25. [ISBN: 0-262-07083-9] "Our goal is to understand the reconstruction of three-dimensional surfaces from two dimensional images at the level of computational theories and algorithms." Concentrates on algorithms "used by the human visual system:" a stereo algorithm to identify points in the images of the right and left eye that correspond to the same location on a surface (using zero-crossings of the convolution of image irradiances with a Laplacian or Gaussian filter), followed by a surface interpolation algorithm to define the unique surface which minimizes the quadratic variation of all surfaces satisfying the stereo algorithm. LAS

Applications (Environment), P. Environmetrics 81: Selected Papers. SIAM, 1981, vii + 259 pp, $18.50 (P). [ISBN: 0-89871-178-9] Seventeen selected papers from an April 1981 SIAM-SIMS-EPA conference in Alexandria, Virginia on the new discipline of environmetrics: mathematics and statistics in the services of the environment. Papers were selected in part for their expository quality; the keynote paper (by J.S. Hunter) especially addresses the special problems of communicating statistical models in plain English. Other papers are case studies of air, water and toxicology systems. LAS

Applications (Physics), S(15-17), L*.** Seventeen Simple Lectures on General Relativity Theory. H.A. Buchdahl. Wiley, 1981, xiii + 174 pp, $23.95. [ISBN: 0-471-09684-9] A philosophically oriented critical survey of general relativity, addressing conceptual problems and recurring perplexities. Intended for those with a "sound knowledge of Euclidean tensor calculus." LAS

Applications (Politics), L, P.** Fair Representation: Meeting the Ideal of One Man, One Vote. Michael L. Balinski, H. Peyton Young. Yale U Pr, 1982, xi + 191 pp. [ISBN: 0-300-02724-9] A popular account of theories of apportionment illustrating both theoretical and historical aspects of the various criteria (single divisor, staying within quota) and corresponding defects (population paradoxes, Alabama paradox). Major results (proved in a lengthy, theoretical appendix): the present method of apportionment is biased in favor of small states; only divisor methods avoid the various population paradoxes; and no method that avoids the population paradoxes will always stay within quota. Balinski and Young, expanding on arguments in their 1976 Monthly article, argue that Webster's method is the fairest compromise method of apportionment. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

1982 Officers and Committees

Mathematical Association Of America
Officers and Committees
March 15, 1982

General Offices: Dolciani Mathematical Center
1529 Eighteenth Street, N.W., Washington, DC 20036
Telephone 202-387-5200

Executive Director: Alfred B. Willcox
Associate Director: Marcia P. Sward
Editorial Director: Raoul Hailpern
Secretary Emeritus: Henry L. Alder

Officers

President, Richard D. Anderson, Louisiana State University (1981-82)
President-elect, Ivan Niven, University of Oregon (1982-)
Past-president, Dorothy L. Bernstein, Brown University (1981-82)
First Vice-president, Ronald L. Graham, Bell Labs (1982-83)
Second Vice-president, William G. Chinn, San Francisco City College (1981-82)
Editor, Paul R. Halmos, Indiana University (1982-1986)
Secretary, David P. Roselle, Virginia Tech (1980-84)
Treasurer, Leonard Gillman, University of Texas (1978-82)

Additional Members of the Board of Governors

Past-president: Dorothy L. Bernstein, Brown University (1981-82)
Elected Members of the Finance Committee: G. Baley Price and Donald L. Kreider
Editor of Mathematics Magazine: Doris J. Schattschneider, Moravian College (1981-85)
Editor of The Two-year College Mathematics Journal: Donald J. Albers, Menlo College (1979-83)

Governors-at-Large:
Susan J. Devlin, Bell Labs (1980-82)
Richard K. Guy, University of Calgary (1980-82)
Herbert S. Wilf, University of Pennsylvania (1981-83)
Martha Zelinka, Weston High School (1981-83)
Peter L. Renz, W. H. Freeman, San Francisco (1982-1984)
Marilyn J. Zweng, University of Iowa (1982-84)

Sectional Governors (July 1, 1979-June 30, 1982):
Kansas, John J. Hutchinson, Wichita State University

Missouri, Troy L. Hicks, University of Missouri, Rolla
New Jersey, Michael I. Aissen, Rutgers University
Northeastern, Anne F. O'Neill, Wheaton College
Ohio, Samuel W. Hahn, Wittenberg University
Pacific Northwest, Ivan Niven, University of Oregon
Seaway, Mabel D. Montgomery, SUC at Buffalo
Southwestern, Gerald S. Rogers, New Mexico State University
Southeastern, John D. Neff, Georgia Tech

Sectional Governors (July 1, 1980-June 30, 1983):
Florida, Beverly L. Brechner, University of Florida
Illinois, Gordon D. Mock, Western Illinois University
Intermountain, C. Edmund Burgess, University of Utah
Iowa, William L. Waltmann, Wartburg College
Louisiana-Mississippi, David E. Cook, University of Mississippi
Maryland-DC-Virginia, Ronald M. Davis, Northern Virginia Community College
Michigan, Delia Koo, Eastern Michigan University
North Central, Dale E. Varberg, Hamline University
Eastern Pennsylvania-Delaware, Gerald J. Porter, University of Pennsylvania
Southern California, John Todd, California Tech
Texas, John T. Mohat, North Texas State University

Sectional Governors (July 1, 1981-June 30, 1984):
Allegheny Mountain, Charles A. Cable, Allegheny College
Indiana, Meyer Jerison, Purdue University
Kentucky, Martha F. Watson, Western Kentucky University
Metropolitan New York, Donald J. Mccarthy, St. John's University
Nebraska, Gerald W. Johnson, University of Nebraska
Northern California, Jean J. Pedersen, University of Santa Clara
Oklahoma-Arkansas, James H. Yates, Central State University
Rocky Mountain, John H. Hodges, University of Colorado
Wisconsin, Norbert J. Kuenzi, University of Wisconsin, Oshkosh

MAA Section Officers

Allegheny Mountain:
Chairman: Franz X. Hiergeist, West Virginia University, Morgantown
First Vice-chairman: Donald M. Platte, Mercyhurst College
Second Vice-chairman: Kathleen A. Taylor, Duquesne University
Coordinator of Student Programs: Barbara Faires, Westminster College
AHSME Regional Exam Coordinators: Nicholas Ford, Pennsylvania State University (W. Pa); I. Dee Peters, University of West Virginia (WV)
Newsletter Editor: Kathleen A. Taylor, Duquesne University
Secretary/Treasurer: John M. Atkins, West Virginia University, Morgantown

Eastern Pennsylvania and Delaware:
Chairman: Bing Wong, Wilkes College
Vice-chairman: David E. Flesner, Gettysburg College
AHSME Regional Exam Coordinators: Willard Baxter, University of Delaware; Albert E. Filano, West Chester State College (E. Pa)
Newsletter Editor: Paul Cochrane, Bloomsburg State College
Secretary/Treasurer: Willard E. Baxter, University of Delaware, Newark

Florida:
Chairman: Robert Gilmer, Florida State University, Tallahassee
Chairman-elect: Charles W. Nelson, University of Florida, Gainesville
Vice-chairman: Marilyn L. Repsher, Jacksonville University
Vice-chairman: Maurice E. Nott, St. Petersburg Junior College
Past Chairman: Edwin Duda, University of Miami, Coral Gables
AHSME Regional Exam Coordinators: William Hoop, American Bankers Assurance Company (FL); Walter E. Mientka, University of Nebraska (Canal Zone); Eugene A. Francis, University of Puerto Rico (PR)
Public Relations: Frank L. Cleaver, University of South Florida, Tampa
Newsletter Editor: Frank L. Cleaver, University of South Florida, Tampa
Secretary/Treasurer: Frank L. Cleaver, University of South Florida, Tampa

Illinois:
Chairman: John S. Haverhals, Bradley University
Chairman-elect: Richard F. Johsonbaugh, Chicago State University
First Vice-chairman: Ronald M. Shelton, Milliken University
Second Vice-chairman: John W. Leduc, Eastern Illinois University
AHSME Regional Exam Coordinator: Gary Tippett, Bradley University
Secretary/Treasurer: Howard C. Saar, 216 Willow Dr. E., Plainfield, Il

Indiana:
Chairman: Marvin G. Mundt, Valparaiso University
Vice-chairman: Carl C. Cowen, Jr., Purdue University
AHSME Regional Exam Coordinator: Stephen Terry, The Nyhart Company
Public Relations: Roger Lautzenheiser, Rose-Hulman Inst of Tech
Newsletter Editor: Roger Lautzenheiser, Rose-Hulman Inst of Tech
Secretary/Treasurer: Richard R. Patterson, Indiana-Purdue University

Intermountain:
Chairman: Richard Tebbs, Southern Utah State College
First Vice-chairman: Richard R. Miller, Weber State College
Second Vice-chairman: Linda C. Hill, Idaho State University, Pocatello
AHSME Regional Exam Coordinator: Ralph V. Skarda, Brigham Young University (UT); Tom Williams, Idaho State University (ID)
Secretary/Treasurer: Lawrence O. Cannon, Utah State University

Iowa:
Chairman: A. M. Fink, Iowa State University
Chairman-elect: Edward T. Hill, Cornell College
AHSME Regional Exam Coordinator: Charles Smith, The Bankers Life
Secretary/Treasurer: James Peake, Iowa State University

Kansas:
Chairman: Sr. Jo Ann Fellin, Benedictine College
Vice-chairman: John W. Carlson, Emporia State University
Associate Chairman for Community Colleges: Louis D. Kottman, Jr., Barton Co. Comm. College
AHSME Regional Exam Coordinator: Steven L. Cooper, Security Benefit Group
Secretary/Treasurer: Robert H. Thompson, Washburn University

Kentucky:
Chairman: Richard M. Davitt, University of Louisville
Chairman-elect: Peter E. Moore, Northern Kentucky University
Vice-chairman: Edgar E. Enochs, University of Kentucky, Lexington
AHSME Regionl Exam Coordinator: William Parker, Jefferson Community College, SW
Secretary/Treasurer: Joe K. Smith, Northern Kentucky University

Louisiana-Mississippi:
Chairman: Paul W. Spikes, Mississippi State University
LA Vice-chairman: Kathleen A. Drude, Northeast Louisiana University
MS Vice-chairman: Ken Johnston, Hinds Junior College
Past Chairman: Adam J. Hulin, University of New Orleans

AHSME Regional Exam Coordinators: Benjamin Mitchell, Louisiana State University (LA); David E. Cook, University of Mississippi (MS)
Newsletter Editor: John L. Tilley, Mississippi State University
Secretary/Treasurer: John L. Tilley, Mississippi State University

Maryland-DC-Virginia:
Chairman: John Schmeelk, Virginia Commonwealth University
Vice-chairman for Programs: Howard L. Penn, US Naval Academy
Vice-chairman for Membership: Ernest L. Mabrey, 1529 Roxanna Road, NW, Washington, DC
Treasurer: Arthur T. Charlesworth, University of Richmond
AHSME Regional Exam Coordinator: Edward D. Bender, J. Sargeant Reynolds Community College
Newsletter Editor: Ernest L. Mabrey, 1529 Roxanna Road, NW, Washington, DC
Secretary: John R. Hanson, James Madison University

Metropolitan New York:
Chairman: Theresa J. Barz, St. John's University
Vice-chairman for Four-Year Colleges: John A. Thorpe, SUNY at Stony Brook
Vice-chairman for Two-Year Colleges: Douglas S. Brown, Nassau Community College
Vice-chairman for High Schools: Cecile M. Cohen, Horace Mann School, Bronx
Treasurer: Robert G. Payton, Adelphi University
AHSME Regional Exam Coordinator: Sandra Pulver, Pace University
Newsletter Editor: William F. Orr, Hofstra University
Secretary: Lily E. Christ, John Jay College of Criminal Justice

Michigan:
Chairman: Harold T. Slaby, Wayne State University
Vice-chairman: David James, University of Michigan, Dearborn
Vice-chairman: Philip H. Mahler, Henry Ford Community College
AHSME Regional Exam Coordinator: Joanne Rankin, Eastern Michigan University
Newsletter Editor: Delano Wegener, Central Michigan University
Secretary/Treasurer: Paul K. Garlick, University of Michigan, Flint

Missouri:
President: A. Glen Haddock, University of Missouri, Rolla
Vice-president: Kenneth W. Lee, Missouri Western State College
Chairman of High School Lecturers: Leonard Palmer, Southeast Missouri State University
Past President: Merry Mcdonald, Northwest Missouri State University
AHSME Regional Exam Coordintor: Alvin Tinsley, Central Missouri State University
Newsletter Editor: Ed W. Huffman, Southwest Missouri State University
Secretary/Treasurer: Jerry Wilkerson, Missouri Western State College

Nebraska:
Chairman: Randall K. Heckman, Kearney State College
Chairman-elect: John K. Karlof, University of Nebraska, Omaha
AHSME Regional Exam Coordinator: Richard Barlow, Kearney State College
Newsletter Editors: H. M. Cox & T. S. Shores, University of Nebraska, Lincoln
Secretary/Treasurer: Henry M. Cox, 1145 N. 44th St., Lincoln, Nebraska

New Jersey:
Chairman: Susan G. Marchand, Kean College
Vice-chairman for Innovation: Dorothy R. Wolff, Caldwell College
Vice-chairman for Speakers: Sr. Stephanie Sloyan, Georgian Court College
Vice-chairman for Two-Year Colleges: Margaret S. Piedem, Somerset Community College
Vice-chairman for High Schools: Francis E. Masat, Glassboro State College
Treasurer: Judith B. Seery, Bell Telephone Labs, Murray Hill
AHSME Regional Exam Coordinator: Robert Mitchell, Glassboro State College
Newsletter Editor: Jean M. Lane, Union College
Secretary: James J. Magliano, Union College

North Central:
President: Roger B. Kirchner, Carleton College
President-elect: Sabra S. Anderson, University of Minnesota, Duluth
Member at Large: James S. Rue, University of North Dakota, Grand Forks
Member at Large: Harlan A. Hewitt, North Hennepin Community College
Past President: Stephen R. Hilding, Gustavus Adolphus College
AHSME Regional Exam Coordinators: Wayne Roberts, Macalester College (MN); John Whitcomb, University of North Dakota (ND); John Barsby, St. John's Ravenscourt (Manitoba); Don Kapoor, University of Regina (Saskatchewan)
Public Relations: Joseph Gallian, University of Minnesota, Duluth
Newsletter Editor: Joseph Gallian, University of Minnesota, Duluth
Secretary/Treasurer: Allan M. Kirch, Macalester College

Northeastern:
Chairman: James E. Ward, Bowdoin College
Two-Year College Rep: Nancy Myers, 67 Highland Ave., Cambridge, MA
AHSME Regional Exam Coordinators Paul Norris, Connecticut Mutual Life Insurance Company (CT); Olin Sawyer, Union Mutual Life Insurance Company (ME); Murray Abramson, Bridgewater State College (MA); Owen R. Koppang, Concord High School (NH & RI); J. S. H. Johansson, University of Vermont (VT); William Crawford, Mt. Allison University (Newfoundland, New Brunswick, Nova Scotia, PE Island)
Newsletter Editor: Eric C. Nummela, New England College
Secretary/Treasurer: Shirley A. Blackett, 97 Eliot Ave., W. Newton, MA

Northern California:
Chairman: Hans Samelson, Stanford University
Vice-chairman: Roy W. Ryden, Humboldt State University
Program Chairman: Carroll O. Wilde, Naval Postgraduate School
AHSME Regional Exam Coordinators: Robert Mcfarland, Alameda County Office of Education (No. CA & NV); Jack Johnson, Brigham Young University (HI, Pacific Islands, Guam)
Newsletter Editor: William G. Chinn, 539 29th Ave., San Francisco, CA
Secretary/Treasurer: Leonard F. Klosinski, University of Santa Clara

Ohio:
Chairman: J. Douglas Faires, Youngstown State University
Chairman-elect: Darrell J. Horwath, John Carroll University
Program Chairman: Alan G. Poorman, Ashland College
Two-Year College Rep: Cynthia P. Yang, Miami University
Past Chairman: Clifford A. Long, Bowling Green State University
AHSME Regional Exam Coordinators Leo J. Schneider, John Carroll University (OH); I. Dee Peters, University of West Virginia (WV)
Public Relations: Susan G. Schneider, John Carroll University
Newsletter Editor: David E. Kullman, Miami University
Secretary/Treasurer: Gus Mavrigian, Youngstown State University

Oklahoma-Arkansas:
Chairman: Robert D. Mcmillan, Oklahoma Christian College

First Vice-chairman: Allan C. Cochran, University of Arkansas, Fayetteville
Second Vice-chairman: Morris L. Marx, University of Oklahoma, Norman
Past Chairman: Andrew D. Coe, Westark Community College
AHSME Regional Exam Coordinators: Thomas Cairns, University of Tulsa (OK); Claude Duplissey, University of Arkansas, Little Rock (AR)
Public Relations: James E. Scroggs, University of Arkansas, Fayetteville
Newsletter Editor: Raymond E. Beasley, Central State University
Secretary/Treasurer: John Jobe, Oklahoma State University

Pacific Northwest:

Chairman: Beryl M. Green, Western Oregon State College
Vice-chairman of Two-Year Colleges: Richard H. Plagge, Highline College
Vice-chairman of Four-Year Colleges: Gail A. Williams, Western Washington University
AHSME Regional Exam Coordinators: Donval R. Simpson, Institute of Arctic Biology (AK); Tom Williams, Idaho State University (ID); William Myers, University of Montana (MT); Michael N. Dyer, University of Oregon (OR); Calvin Long, Washington State University (WA); Harold L. Lampkin, University of Calgary (Alberta); Leon Bowden, University of Victoria (British Columbia, Yukon, Northwest Territories)
Newsletter Editor: Sue Kaplan, Western Washington University
Secretary/Treasurer: Sam L. Dunn, Seattle Pacific University

Rocky Mountain:

Chairman: John P. Gill, University of Southern Colorado
Chairman-elect: George S. Donovan, Metropolitan State College
Vice-chairman: Aubrey Owen, Community College of Denver
Program Chairman: Gale Nash, Western State College
AHSME Regional Exam Coordinators: Bob Vunovich, University of Southern Colorado (CO & WY); William Myers, University of Montana (MT)
Newsletter Editor: David W. Ballew, SD School of Mines & Tech
Secretary/Treasurer: David W. Ballew, SD School of Mines & Tech

Seaway:

Chairperson: Kenneth D. Magill, SUNY Center at Buffalo
First Vice-chairperson: Jack E. Graver, Syracuse University
Second Vice-chairperson: Robert C. Knapp, Herkimer County Community College
Past Chairperson: Howard E. Bell, Brock University
AHSME Regional Exam Coordinators: Dennis Martin, SUNY College at Brockport (Upper NY); Louis F. Scholl, Trocaire College (Western NY); F. A. Sherk, University of Toronto (Ontario); Saverio Mirarchi, Laval Catholic High School (Quebec)
Newsletter Editor: Donald Trasher, SUNY College at Geneseo
Secretary/Treasurer: Donald Trasher, SUNY College at Geneseo

Southeastern:

Chairman: James H. Carruth, University of Tennessee, Knoxville
Chairman-elect: Mary F. Neff, Emory University
Vice-chairman: Catherine C. Aust, Clayton Junior College
Lecturer: C. Ray Wylie, Jr., Route 7, Merritt View Terr, Greenville, SC
AHSME Regional Exam Coordinators: James O'Neal, Troy State Univ (AL); Jan List Boal, Georgia State University (GA); Hal B. Phillips, Jr., Jefferson Stand Life Ins Co (NC); Stanley Lukawecki, Clemson University (SC); David Collings, National Life & Accident Ins Co (TN)
Newsletter Editor: Dewayne S. Nymann, University of Tennessee, Chattanooga
Secretary/Treasurer: Ivey C. Gentry, Wake Forest University

Southern California:

Chairman: John Greever, Harvey Mudd College
First Vice-chairman: Judith Grabiner, California State University, Dominguez Hills
Second Vice-chairman: Robert F. Brown, University of California, Los Angeles
Program Chairman: Solomon W. Golomb, University of Southern California
Past Chairman: Adil M. Yaqub, University of California, Santa Barbara
AHSME Regional Exam Coordinator: Dwight O. Coblentz, Department of Ed
Newsletter Editor: J. Myron Hood, California Polytech State University
Secretary/Treasurer: J. Myron Hood, California Polytech State University

Southwestern:

Chairman: William Y. Velez, University of Arizona, Tucson
Vice-chairman: Ralph W. Ball, New Mexico Institute of Mining & Tech
AHSME Regional Exam Coordinator: D. R. Arterburn, New Mexico Institute of Mining & Tech (AZ)
Secretary/Treasurer: Edward D. Gaughan, New Mexico State University, Las Cruces

Texas:

Chairman: Robert L. Tennison, University of Texas, Arlington
Chairman-elect: John Ed Allen, North Texas State University
Level I Director: Vivian A. Dennis, Eastfield College
Level II Director: R. Vic Morgan, Sul Ross State University
Level III Director: Bennie B. Williams, University of Texas, Arlington
Director at Large: Landon Colquitt, Texas Christian University
Arrangements Chair: Richard A. Alo, Lamar University
Arrangements Chair-elect: Melvin R. Hagan, North Texas State University
Past Chairman: Bill D. Anderson, East Texas State University
AHSME Regional Exam Coordinator: James R. Boone, Texas A & M University
Public Relations: James C. Bradford, Abilene Christian University
Newsletter Editor: Glen E. Mattingly, Sam Houston State University
Secretary/Treasurer: Glen E. Mattingly, Sam Houston State University

Wisconsin:

Chairman: Karl Beres, Ripon College
Chairman-elect: Lois B. Brualdi, Edgewood College
Historian: Harvey C. Mckenzie, University of Wisconsin, Oshkosh
Past Chairman: Anthony E. Barkauskas, University of Wisconsin, La Crosse
AHSME Regional Exam Coordinator: Arthur Brownell, Jr., University of Wisconsin, La Crosse
Newsletter Editor: Robert W. Langer, University of Wisconsin, Eau Claire
Secretary/Treasurer: Paul J. Campbell, Beloit College

Committees of the Association

Terms of members expire, except where otherwise noted, at the Annual Meeting in January following the last year of service listed below. For temporary committees, no terms are listed since they are automatically discharged at the expiration of the President's term of office, which is the Annual Meeting in January, 1983.

Executive Committee:

Richard D. Anderson, Chairman (1981-82); William G. Chinn (1982-82); Paul R. Halmos (1982-86); Ronald L. Graham (1982-83); Leonard Gillman (1978-82); Ivan Niven (1982-83); David P. Roselle (1980-84), all are ex-officio.

<u>Finance Committee</u>:
Richard D. Anderson, Chairman (1981-82), ex-officio; Dorothy L. Bernstein (1981-82), ex-officio; Leonard Gillman (1978-82), ex-officio; Donald L. Kreider (1978-85); G. Baley Price (1980-83); David P. Roselle (1980-84), ex-officio.

<u>Audit and Budget Committee</u>: G. Baley Price, Chairman (1980-83); Donald L. Kreider, (1982-85).

<u>Investment Committee</u>: Leonard Gillman, Chairman (1978-82), ex-officio; Henry L. Alder (1982-83); Harley Flanders (1982-83); David P. Roselle (1980-84), ex-officio; Frederic T. Metcalf (1981-82); James M. Vaughn, Jr. (1981-83).

<u>Staff and Services Committee</u>: Leonard Gillman, Chairman (1979-82); Richard D. Anderson (1981-82), both ex-officio.

<u>Advisory Committee on The Archives Of American Mathematics</u>:
Judith V. Grabiner, Chairman (1979-82); Leonard Gillman (1979-82); Robert E. Greenwood (1979-82); Albert C. Lewis, ex-officio; Lucille E. Whyburn (1981-83).

<u>Committee on Advisement and Personnel</u>:
Bernice L. Auslander, Chairman (1980-82); Jane M. Day, (1981-83); Gordon Raisbeck (1980-81); Martha K. Smith (1980-81); William A. Stannard (1981-83); James W. Vick (1980-82).

<u>Committee on Corporate Members</u>:
Richard A. Moynihan (1981-82); Clayton V. Aucoin (1981-83); Jeremiah J. Lyons (1981-83); Stephen H. Quigley (1980-82); James M. Vaughn (1980-82); Alfred B. Willcox (1980-82), ex-officio.

<u>Committee on Earle Raymond Hedrick Lectures</u>:
Mary E. Rudin, Chairman (1980-82); George E. Andrews (1981-82); Daniel Gorenstein (1982-84).

<u>Committee on High School Contests</u>:
Stephen B. Maurer, Chairman (1981-83); George Berzsenyi (1980-82); Thomas R. Butts (1982-84); Richard T. Driver (1981-83); Samuel L. Greitzer (1981-83); Mary Jones (1982-83); Murray Klamkin (1981-83); Walter E. Mientka, Director (1979-82); Richard S. Pieters (1981-83); Stanley Rabinowitz (1979-84); Alexander Scheitlin (1980-82); Leo J. Schneider (1980-82); John H. Staib (1982-84); Nura Turner (1980-82); Margaret Wilkinson (1982-85); Melvin R. Woodard (1981-82); Martha Zelinka (1980-82).

<u>Advisory Panel of the Committee on High School Contests</u>: James D. Bristol; Michael W. Ecker; Paul Foerester; Hubert L. Hunzeker; Irwin W. Kaufman; Katherine P. Layton; Nathan I. Mendelsohn; Attila Mate; Harold B. Reiter; A. W. Roberts; Eugene Veklerov.

<u>Subcommittee on the USA Mathematical Olympiad</u>: Samuel L. Greitzer, Chairman (1980-82); George Berzsenyi (1980-82); Mary H. Jones (1982-83); Murray S. Klamkin (1980-82); Andrew Liu (1981-83); Stephen B. Maurer (1981-83), ex-officio; Walter E. Mientka (1981-82), ex-officio; Cecil C. Rousseau (1982-84); John H. Staib (1982-84), ex-officio.

<u>Advisor to the Olympiad Awards Ceremony</u>. Nura D. Turner (1979-82).

<u>Committee on National Awards and Public Representation</u>:
Peter J. Hilton, Chairman (1980-82); Richard D. Anderson (1981-82), ex-officio; Lester H. Lange (1981-83); Robert Wilson (1980-82).

<u>Committee on Publications</u>:
Edwin F. Beckenbach, Chairman (1980-82); Donald J. Albers (1979-83), ex-officio; Lida K. Barrett (1981-83); Daniel T. Finkbeiner, II (1981-83); Ross L. Finney (1980-82); Leonard Gillman (1978-82), ex-officio; Robert Gilmer (1980-82); Paul R. Halmos (1982-86), ex-officio; Ross A. Honsberger (1980-82); William F. Lucas (1982-84); Barbara L. Osofsky (1981-83); Warren Page (1981-83); Doris Schattschneider (1981-85), ex-officio; Alan C. Tucker (1979-85).

<u>Subcommittee on Basic Library Lists</u>: Daniel T. Finkbeiner, II, Chairman (1981-83); Donald W. Bushaw (1980-81); H. E. Hall (1980-82); Robert H. Mcdowell (1980-82); Frank L. Wolf (1980-82).

<u>Subcommittee on Carus Monographs</u>: Barbara L. Osfsky, Chairman (1981-83); Daniel T. Finkbeiner, II (1981-83); Robert Gilmer (1980-82).

<u>Subcommittee on Carl B. Allendoerfer Awards</u>: Thomas W. Tucker, Chairman (1980-82); Edwin F. Beckenbach (1980-82); J. A. Seebach (1982-84).

<u>Subcommittee on Dolciani Mathematical Expositions</u>: Ross A. Honsberger, Chairman (1981-83); Gerald L. Alexanderson (1981-82); Joseph Malkevitch (1981-83); Kenneth R. Rebman (1981-82).

<u>Subcommittee on George Polya Awards</u>: Warren Page, Chairman (1980-82); Edwin F. Beckenbach (1980-82); Ronald M. Davis (1982-84).

<u>Subcommittee on Lester R. Ford Awards</u>: Peter L. Duren, Chairman (1981-82); Edwin F. Beckenbach (1980-82); David Gale (1982-84).

<u>Subcommittee on MAA Studies in Mathematics</u>: Alan C. Tucker, Chairman (1980-82); Donald W. Anderson (1981-83); Douglas B. West (1982-84).

<u>Subcommittee on Miscellaneous Publications</u>: Edwin F. Beckenbach, Chairman (1980-82); Leonard Gillman (1980-82), ex-officio; David P. Roselle (1980-84), ex-officio.

<u>Subcommittee on New Mathematical Library</u>: Ivan Niven, Chairman (1981-83); William G. Chinn (1980-82); Basil Gordon (1980-82); Anneli Lax, ex-officio; Max M. Schiffer (1979-81).

<u>Editor of the New Mathematical Library</u>: Anneli Lax.

<u>Committee on Articulation</u>:
Billy E. Rhoades, Chairman (1982-84); Philip C. Curtis, Jr. (1982-84); Chancey Jones (1982-84); Joan P. Leitzel (1982-84); Vincent O'Connor (1982-84); Barbara Wickless (1982-84); Gail S. Young (1982-84).

<u>Committee on Mini-courses</u>:
John O. Riedl, Chairman (1982-84); Frederick Hoffman (1982-84); Warren Page (1982-85); Jean Pedersen (1981-83).

<u>Committee on National Meeting Programs</u>:
C. E. Burgess, Chairman (1982-84); Ronald L. Graham (1981-83); Peter A. Lindstrom (1982-85); David P. Roselle (1980-84), ex-officio.

<u>Committee on Secondary School Lectures</u>:
Donald B. Small, Chairman (1979-83); Edward Z. Andalafte (1980-82); Marjorie M. Enneking (1979-82); Gloria F. Gilmer (1980-82); John M. Jobe (1981-83); William K. Mcnabb (1980-82).

<u>Committee on Sections</u>:
Samuel W. Hahn, Chairman (1981-84); Yousef Alavi (1980-83); David W. Ballew (1981-84); Barbara T. Faires (1982-85); John D. Neff (1979-82); Kenneth R. Rebman (1981-84); Alfred B. Willcox, ex-officio.

<u>Committee on Special Funds of the Association</u>:
G. Baley Price, Chairman (1981-83); Joshua Barlaz (1980-82); Lida K. Barrett (1981-83); Leonard Gillman (1980-82), ex-officio; John M. Jobe (1982-84); Eileen Poiani (1982-84); Richard M. Schori (1982-84).

<u>Committee on the Award For Distinguished Service to Mathematics</u>:
Henry L. Alder, Chairman (1980-82); Ralph P. Boas (1981-83); Eileen L. Poiani (1982-84).

<u>Committee on the Chauvenet Prize</u>:
Philip J. Davis (1980-82); David Eisenbud (1981-83).

<u>Committee on the Exchange of Information on Mathematics</u>:
Yousef Alavi, Chairman (1981-83); Donald M. Hill (1981-83); James R. Leitzel (1979-82); Jean J.

Pedersen (1981-83); Lynn A. Steen (1979-82); Marion Wetzel (1980-82).

Committee on the Membership of the Association:

Dorothy L. Bernstein, Chairman (1981-83); Richard D. Anderson (1980-82), ex-officio; Ronald M. Davis (1980-82); Leonard Gillman (1978-82); Samuel W. Hahn (1982-84); David P. Roselle (1980-84), ex-officio; Sue H. Whitesides (1981-83); Paul R. Zuckerman (1980-82).

Committee on Placement Examinations:

Bernard L. Madison, Chairman (1980-82); Richard H. Prosl, Chairman (1982-84); Thomas A. Carnevale (1980-82); John W. Kenelly (1980-82); Billie A. Rice (1981-83); Thomas W. Tucker (1982-84).

Committee on the Putnam Prize Competition:

William Firey, Chairman (1980-82); Gerald L. Alexanderson, Associate Director (1980-86); Abraham P. Hillman, Associate Director (1978-86); Leonard F. Klosinski (1980-86); Douglas A. Hensley (1981-83); Melvin Hochster (1982-84).

Committee on the Teaching of Undergraduate Mathematics:

James W. Vick, Chairman (1980-82); Donald W. Bushaw (1980-82); Ronald M. Davis (1982-84); Gloria F. Gilmer (1981-83); David I. Schneider (1981-83); Alan H. Schoenfeld (1979-82); Martha J. Siegel (1981-83); James E. Ward, III (1982-84); June P. Wood (1981-83).

Committee on Two-Year Colleges:

John D. Bradburn, Chairman (1981-83); Donald J. Albers (1980-82); Thomas A. Carnevale (1979-83); Betsy Darden-Smith (1982-84); Ann D. Holley (1982-84); J. C. Moss (1979-83); Warren Page (1980-82); Richard H. Plagge (1980-82); Elaine L. Tatham (1981-83); William R. Wilson (1982-84).

Committee on the Undergraduate Program in Mathematics:

Donald W. Bushaw, Chairman (1979-82); Gerald L. Alexanderson (1980-82); Richard D. Anderson, ex-officio (1980-82); Winifred A. Asprey (1980-82); Betty A. Case (1982-84); James A. Donaldson (1982-84); Jerome A. Goldstein (1980-84); Joan P. Leitzel (1982-84); Stephen B. Maurer (1981-83); Bruce E. Meserve (1980-82); Thomas G. Proctor (1982-84); Billy E. Rhoades (1981-83); Donald B. Small (1980-82).

Committee on Visiting Lecturers and Consultants:

Malcolm W. Pownall, Chairman (1980-82); Theodore S. Chihara (1981-83); Vivian Y. Kraines (1981-83); Jon M. Laible (1980-82); Eugene R. Seelbach (1979-84).

Committee on World War II History:

G. Baley Price, Chairman (1981-83); Churchill Eisenhart (1981-83); Carroll V. Newsom (1981-83); Mina S. Rees (1981-83); J. B. Rosser (1981-83); Raymond L. Wilder (1981-83).

Joint Committees

Joint Committee on Employment Opportunities:

Wilfred E. Barnes, Chairman (1984, MAA); James W. Daniel (1981-83, SIAM); Roger A. Horn (1978-82, AMS); Calvin Long (1981-83, MAA); Ervin Y. Rodin (1978-82, SIAM); Donald C. Rung (1981-82, AMS).

Joint Committee on Women in Mathematics:

Mary W. Gray, Chairman (1981-83); Etta Z. Falconer (1982, AMS); Pamela Cook-Ioannidis (1983, SIAM); Carol B. LaCampagne (1984, MAA); Edith H. Luchins (1982, MAA); Katherine Pedersen (1983, NCTM); Joel Schneider (1983, NCTM); Barbara Searle 1983, NCTM); Gail A. Williams (1983, MAA).

Joint Meetings Committee:

William J. Leveque, Chrirman; Everett Pitcher; David P. Roselle; Alfred B. Willcox, all ex-officio.

AMS-MAA-SIAM Joint Projects Committee for Mathematics:

C. Edmund Burgess (1983); Wendell H. Fleming (1982); Shirley A. Hill (1983); Gottfried E. Noether (1982); Seymour V. Parter (1982). (Terms expire September 30.)

CEEB-MAA Committee on Mutual Concerns:

Henry L. Alder (1982, MAA); Richard D. Anderson (1983, MAA); Lida K. Barrett (CEEB); Chancey O. Jones (CEEB); Donald L. Kreider (CEEB); David P. Roselle, Co-chairman (1984, MAA); Howard E. Taylor, Co-chairman (CEEB).

Editorial Boards of the Association

American Mathematical Monthly: (All terms expire December 31, 1986)

Editor: Paul R. Halmos

Associate Editors: Sheldon Axler, Thomas F. Banchoff, Ralph P. Boas, David Borwein, Joel L. Brenner, Vladimir Drobot, Allan L. Edmonds, John H. Ewing, Richard K. Guy, Raoul Hailpern, Kenneth R. Rebman, Seymour Schuster, J. A. Seebach, Lynn A. Steen, Robert F. Wardrop, Mary R. Wardrop, Herbert S. Wilf.

Mathematics Magazine: (All terms expire December 31, 1985)

Editor: Doris L. Schattschneider

Associate Editors: Edward J. Barbeau, John A. Beidler, Paul J. Campbell, Underwood Dudley, Gerald A. Edgar, Joseph A. Gallian, Judith V. Grabiner, Raoul Hailpern, Pierre J. Malraison, Joseph Malkevitch, William A. Mcworter, Jr., Leroy F. Meyers, Jean J. Pedersen, Gordon Raisbeck, Ian Richards, Eric S. Rosenthal, Doris J. Schattschneider, Dan Shapiro, David A. Smith.

Two-Year College Mathematics Journal: (All terms expire December 31, 1983)

Editor: Donald J. Albers

Associate Editors: Gerald L. Alexanderson, Glenn D. Allinger, Anthony Barcellos, William G. Chinn, Ronald M. Davis, Howard W. Eves, Stanley Friedlander, Thomas M. Green, Samuel A. Greenspan, Raoul Hailpern, Ross A. Honsberger, Harold R. Jacobs, Erwin Just, Bruce W. King, Norman F. Ladd, Roland H. Lamberson, William W. Leonard, Peter A. Lindstrom, David E. Logothetti, Nancy Myers, Warren Page, George Polya, Edward B. Wright.

Newsletter Editorial Committee: (All terms expire December 31, 1985)

J. Arthur Seebach, Jr.; Co-chairman, Lynn A. Steen, Co-chairman; Yousef Alavi, ex-officio; Henry L. Alder; Philip M. Cheifetz; Ronald M. Davis; Katherine P. Layton; Eileen L. Poiani; Kenneth R. Rebman; David P. Roselle, ex-officio; Marcia P. Sward, ex-officio; Alan C. Tucker.

Representatives Of The Association

On Sections of AAAS:

Section A: Lowell J. Paige (1981-83)
Section Q: Joan P. Leitzel (1981-83)
Section U: Fred Mosteller (1981-83)
Section X: Alfred B. Willcox (1981-83)

On UMAP Editorial Board: Anne F. O'Neill (1981-82); Peter A. Lindstrom (1981-82); Philip D. Straffin (1982-84).

On the Council of the Conference Board of the Mathematical Sciences: Richard D. Anderson.

On Discipline Committee for Mathematics of CEEB: Donald L. Kreider.

On the Governing Council of Mu Alpha Theta: Katherine P. Layton (1982-84).

On the U. S. Commission on Mathematical Instruction: Leon A. Henkin (1978-82); Eileen L. Poiani (1981-83).

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

ON THE NONSOLVABILITY OF THE GENERAL POLYNOMIAL

RAYMOND G. AYOUB
Department of Mathematics, Pennsylvania State University, 203 McAllister Building, University Park, PA 16802

1. Introduction. The general polynomial of degree $n \geqslant 5$ is not solvable by radicals. This is a celebrated theorem which is generally taught in courses in Galois theory. It is deduced from the relation between the solvability of a polynomial by radicals and the solvability of the Galois group of its splitting field. The Galois group of the general polynomial is S_n, the symmetric group on n symbols and, as this is not a solvable group for $n \geqslant 5$, the result follows.

The theorem was first stated in 1799 by P. Ruffini [**1**] and was proved by him, subject to an assumption that he did not adequately prove, and that was not in essence established until 1826 by N. Abel [**2**]. The arguments used by these two were shortlived, since they were quickly overshadowed by the ideas of Galois.

The reduction of the solvability of a polynomial to that of its Galois group is somewhat lengthy, requiring as it does the notion of solvability of a group and the excursion to Kummer fields. The point of the following proof is to bypass this reduction and go to the specific polynomial under consideration. While it is possible to avoid basic notions of field theory, there is no special merit in doing so and hence we assume a knowledge of the most elementary properties of fields.

2. Some Definitions and Background.

DEFINITION. By a radical extension of a field K, we shall mean a field L for which there is a $\lambda \in L$, $\lambda \notin K$, and a prime p, such that $\lambda^p \in K$ and $L = K(\lambda)$.

DEFINITION. By a radical chain from K to L, we mean a chain

$$K = K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_i \subset \cdots \subset K_m = L$$

where K_i is radical over K_{i-1}. We shall call m the length of the chain.

DEFINITION. Let $f(x) \in K[x]$ and let F be its splitting field. $f(x)$ is said to be solvable by radicals if there exists a field L such that $F \subset L$ and there is a radical chain from K to L.

The concept of solvability by radicals was not greatly elaborated upon by early writers. Let Q be the field of rational numbers and

$$h(x) = x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n \in Q[x].$$

What these writers were looking for was a "formula" for the roots in terms of $a_1, a_2, \ldots, a_n$; the formula was to involve only the operations of addition, subtraction, multiplication, division *and* the extraction of roots. The formula was to be applicable to any polynomial of degree n in $Q[x]$ so that, given any polynomial, the coefficients were to be inserted in the formula and out would come the roots, exactly as in the known cases for $n = 2$, 3 and 4. Thus, these writers were dealing tacitly with the "general" polynomial of degree n to which we shall come shortly. The existence of a formula is tantamount to the definition given above for solvability by radicals, and we shall use the above definition from now on.

Let $x_1, x_2, \ldots, x_n$ be indeterminates (i.e., algebraically independent elements over Q) and let $P = Q(x_1, \ldots, x_n)$ be the field of rational functions in $x_1, \ldots, x_n$ over Q.

DEFINITION. Let K be a field and $L = K(x_1,\dots,x_n)$. By the general polynomial of degree n, over K, we mean the polynomial

$$f(x) = x^n + x_1x^{n-1} + \cdots + x_{n-1}x + x_n \in L[x].$$

What we wish to show is that if $n \geqslant 5$ and $K = Q$, then $f(x)$ is not solvable by radicals (i.e., there can be no "formula" for the roots).

If $f(x)$ is solvable by radicals, then a fortiori every $h(x) \in Q[x]$ is solvable by radicals. But, a priori, it may be that every $h(x) \in Q[x]$ is solvable by radicals but that $f(x) \in P[x]$ is not. We now know that this is not the case. For every n, there exists $h(x) \in Q[x]$ of degree n such that $h(x)$ is not solvable by radicals. See, e.g., [3]. We need dwell no further on this interesting point.

As we wish to show that $f(x)$ is not solvable by radicals, then it suffices to show that it is not solvable over some extension field of P. We therefore enlarge the base field Q by adjoining to Q a primitive $n!$th root of unity ρ and let $H = Q(\rho)$. Let $E = H(x_1,\dots,x_n) = Q(\rho)(x_1,\dots,x_n) = P(\rho)$ and let $F = E(\xi_1, \xi_2,\dots,\xi_n)$ be the splitting field over E; i.e., $\xi_1, \xi_2,\dots,\xi_n$ are the roots of $f(x)$. We require some known results, whose proofs we give for the sake of completeness.

LEMMA. *The roots $\xi_1,\dots,\xi_n$ are algebraically independent over $Q(\rho)$.*

Proof. Note that $F = H(\xi_1,\dots,\xi_n)$ since $x_i \in H(\xi_1,\dots,\xi_n)$ $(i = 1,2,\dots,n)$. Moreover if K is any field and $y_1,\dots,y_n$ are indeterminates over K, then any permutation τ of $y_1,\dots,y_n$ induces an automorphism of $K(y_1,\dots,y_n)$ that leaves K fixed.

Suppose now that $g(y_1,\dots,y_n) = g(y)$ $(y = (y_1,\dots,y_n))$ is a nontrivial polynomial over H for which $g(\xi) = g(\xi_1,\dots,\xi_n) = 0$. Let

$$h(y) = \prod_{\tau \in S_n} g^\tau(y)$$

where $g^\tau(y) = g(y_{\tau(1)},\dots,y_{\tau(n)})$. Then $h(y)$ is symmetric in $y_1,\dots,y_n$ and hence is a polynomial $k(\sigma_1,\dots,\sigma_n) = k(\sigma)$ in the elementary symmetric functions σ_i of $y_1,\dots,y_n$, and its coefficients are in H:

$$h(y) = k(\sigma).$$

If $k(\sigma) = 0$, then some factor $g^\tau(y) = 0$, which implies that $g(y) = 0$, since τ is an automorphism. This contradicts the assumption on $g(y)$. On the other hand, if $y_1,\dots,y_n$ are replaced by $\xi_1,\dots,\xi_n$, then σ_i are replaced by $(-1)^i x_i$ and hence

$$0 = h(\xi) = k(x_1,\dots,x_n).$$

As k is a nontrivial polynomial, this is a contradiction.

COROLLARY 1. *Any permutation of $\xi_1,\dots,\xi_n$ is an automorphism of F leaving E fixed.*

COROLLARY 2. *The polynomial $f(x)$ is irreducible over E. This follows at once from Corollary* 1.

The proof of the main theorem is in two parts. The first part shows that if $f(x)$ is solvable by radicals, we need not leave the field F. The reader will contrast this with the case of a cubic equation over Q, all of whose roots are real. The formula for the roots requires the use of cube roots of unity. The second part shows that, for a radical extension, the behavior of the automorphisms induced by permutations is not arbitrary but is narrowly prescribed.

3. Abel's Theorem. With the notation above, we have

THEOREM A. *If $f(x)$ is solvable by radicals, then there is a radical chain from E to F.*

The idea of the proof is this: Assuming that F is contained in a radical extension K, then we start at the top of the chain joining E to K and show successively that each can be generated over the preceding by an element of F. This induction is carried out with the help of a

LEMMA. *Let H be a field containing the pth roots of unity with p a prime. Let M be a finite extension of H. Let $L = M(\lambda)$, $\lambda^p = \nu \in M$, $\lambda \notin M$ and let F be normal over H. If there exists an element $\alpha \in F \cap L$, $\alpha \notin M$, then there exists an element $\mu \in F \cap L$ such that $M(\mu) = M(\lambda) = L$ with $\mu^p \in M$.*

(All fields are assumed to be contained in a common one.)

Proof. Since $\alpha \in L$, α is a linear combination of powers of λ with coefficients in M. Omitting the terms with zero coefficient, we have

$$\alpha = a_0\lambda^{t_0} + a_1\lambda^{t_1} + \cdots + a_r\lambda^{t_r}$$

with $0 \leq t_i < p$, $a_i \neq 0$, $a_i \in M$. Since $\alpha \notin M$, some t_j is not zero. For simplicity, let $t_j = t$.

There are automorphisms σ_k of L over M such that $\sigma_k(\lambda) = \eta^k\lambda$, where η is a primitive pth root of unity. Hence $\alpha = \alpha^{(1)}$ is taken by σ_k to $\alpha^{(n_k)}$, a conjugate of α, which by the normality of F lies in F. Therefore,

$$\alpha^{(n_k)} = \\ = \eta^{kt_0}(a_0\lambda^{t_0}) + \cdots + \eta^{kt_r}(a_r\lambda^{t_r}) \quad (k = 0,\ldots,r).$$

We now view these as a system of equations in $a_i\lambda^{t_i}$. The determinant of the system is the Vandermonde $|\eta^{it_j}| \neq 0$. We therefore solve for $a_t\lambda^t$, which we denote by μ. Since $\eta \in F$, we conclude that $\mu \in F$. Moreover,

$$\mu^p = a_t^p(\lambda^t)^p = a_t^p(\lambda^p)^t = a_t^p\nu^t \in M$$

and $M(\mu) \subset M(\lambda)$. Finally, since $(t, p) = 1$, we choose a and b such that $at + bp = 1$; then

$$\lambda = \lambda^{at+bp} = (\lambda^t)^a(\lambda^p)^b = \frac{\nu^b}{a_t^a}\mu^a \in M(\mu).$$

Thus $M(\mu) = M(\lambda)$ as required.

Proof of Theorem A. Abel's original proof contains a gap that appears to have been repeated by later writers. The following is, therefore, a modification and rectification of his proof. The basic idea however, remains his.

By assumption, there exists a field K, $F \subset K$ and a radical chain from E to K.

We wish to apply the lemma successively to each of the steps in this chain. In attempting to do so, Abel fails to account for an alternative which might arise. This is the gap we have referred to. This difficulty is circumvented by the following construction:

Among the radical chains from E to K, choose those of minimal length k; and among these choose a chain

$$E = E_0 \subset E_1 \subset \cdots \subset E_r = F_0 \subset F_1 \subset \cdots \subset F_s = K$$

with $r + s = k$ satisfying the following properties:

(i) $E_i = E_{i-1}(\nu_i), \nu_i \in F, \nu_i^{q_i} \in E_{i-1} \quad (i = 1,2,\ldots r)$

(ii) $F_i = F_{i-1}(\lambda_i), \lambda_i^{p_i} \in F_{i-1} \quad (i = 1,2,\ldots s)$

(iii) r is maximal with respect to properties (i), (ii).

If $r = k$, then $s = 0$, (ii) is vacuous, and the theorem is proved. Assume therefore that $s > 0$. Then the set of integers m for which

$$F_m = F_{m-1}(\mu_m), \mu_m \in F, \mu_m^{p_m} \in F_{m-1} \tag{1}$$

is nonempty; for we have $F \subset F_s$, $F \not\subset F_{s-1}$ (otherwise the minimality of k is contradicted). Using the lemma with any $\alpha \in F$, $\alpha \notin F_{s-1}$, $H = E$, $M = F_{s-1}$, $L = F_s$, we find $\mu_s \in F$, such that $F_s = F_{s-1}(\mu_s)$, $\mu_s^{p_s} \in F_{s-1}$. Let m be the least integer satisfying (1); then we claim that

$$\mu_m^{p_m} \in F_0. \tag{2}$$

If not, choose $q > 0$ such that $\mu_m^{p_m} \in F_q$, $\notin F_{q-1}$. Applying the lemma with $\alpha = \mu_m^{p_m}$, we get $\mu_q \in F$, $\mu_q^{p_q} \in F_{q-1}$, $F_q = F_{q-1}(\mu_q)$, with $q < m$ contradicting the choice of m.

Since by (2), $F_0(\mu_m)$ is radical over F_0, the chain

$$\begin{aligned} E_0 \subset E_1 \subset \cdots \subset E_r = F_0 \subset F_0(\mu_m) \subset F_1(\mu_m) \subset \cdots \subset F_{m-1}(\mu_m) \\ = F_m \subset F_{m+1} \subset \cdots \subset F_s = K \end{aligned}$$

is a radical chain of length $\leqslant k$. If the length is $< k$, then the minimality of k is violated, while if the length of the chain is k, the maximality of r is contradicted. The proof is thus complete.

4. Ruffini's Theorem. Ruffini recognized the need to prove Theorem A but did not succeed in giving adequate justification. Assuming the validity of this result, he then gave several proofs of the nonsolvability. We choose the simplest which is based on

THEOREM R. *Let p be a prime and let $A \in F$. If $A^p = B$ and B is invariant under the automorphisms induced by the* 3 *cycles* (1 2 3), (2 3 4), (3 4 5), *then so is A.*

Proof. It is at this stage that we use, in an essential way, the condition $n \geqslant 5$. Let $\sigma_1, \sigma_2, \sigma_3$ be the automorphisms induced by $a_1 = (1\,2\,3)$, $a_2 = (2\,3\,4)$, $a_3 = (3\,4\,5)$. We have

$$a_1 a_2 = (1\,3)(2\,4),\ a_2 a_3 = (2\,4)(3\,5),\ a_1 a_2 a_3 = (1\,4\,2\,5\,3).$$

Therefore

$$(\sigma_2\sigma_1)^2 = (\sigma_3\sigma_2)^2 = (\sigma_3\sigma_2\sigma_1)^5 = 1. \tag{1}$$

We also have

$$\sigma_1^3 = \sigma_2^3 = \sigma_3^3 = 1.$$

Now for $i = 1, 2, 3$,

$$(\sigma_i(A))^p = \sigma_i(A^p) = \sigma_i B = B$$

by assumption; hence

$$\sigma_i A = \zeta_i A \qquad (i = 1, 2, 3) \tag{2}$$

where ζ_i is a pth root of unity. From (1), we get

$$(\zeta_1\zeta_2)^2 = (\zeta_2\zeta_3)^2 = (\zeta_1\zeta_2\zeta_3)^5 = \zeta_1^3 = \zeta_2^3 = \zeta_3^3 = 1.$$

These relations give $\zeta_1\zeta_2 = \zeta_2\zeta_3 = \zeta_1\zeta_2\zeta_3 = 1$ and therefore $\zeta_1 = \zeta_2 = \zeta_3 = 1$ as required.

COROLLARY. *Suppose that $M \subset L \subset F$, and that L is radical over M. If M is left fixed by $\sigma_1, \sigma_2, \sigma_3$, then so is L.*

5. Proof of the Theorem. We restate the result.

THEOREM. *If $n \geqslant 5$ and $f(x)$ is the general polynomial of degree n over $Q(\rho)$, then $f(x)$ is not solvable by radicals.*

Proof. As noted above, if we prove that $f(x)$ is not solvable by radicals over $Q(\rho)(x_1, \ldots, x_n)$, then a fortiori, it is not solvable over $Q(x_1, \ldots, x_n)$.

By Theorem A, there exists a radical chain

$$E = E_0 \subset E_1 \subset \cdots \subset E_{i-1} \subset E_i \subset \cdots \subset E_k = F$$

with

$$\nu_i = \mu_i^{p_i} \in E_{i-1}, \quad E_i = E_{i-1}(\mu_i) \quad \text{and} \quad p_i \text{ prime } (i = 1, 2, \ldots k).$$

By the corollary to Theorem R, if E_{i-1} is left fixed by σ_j $(j = 1, 2, 3)$ then so is E_i; E_0, however,

is left fixed and so by induction F must be left fixed. This, however, is a contradiction since σ_1 takes ξ_1 to ξ_2.

References

1. N. Abel, Collected Works, vol. 1, Christiania Imprimerie d'Grondahl, 1881, pp. 28–34; pp. 66–87.
2. P. Ruffini, Collected Works, vol. 1, Math. Circle of Palermo, Tipografia Matematica, Palermo, Italy, 1950, pp. 159–186.
3. C. R. Hadlock, Field Theory and its Classical Problems, Chapter 4, Carus Math. Monograph No. 19, Math. Assoc. of Amer., Washington, D. C., 1978.

CENTRAL FUNCTIONS AND FIXED POINTS

DAN KALMAN
Department of Mathematics, University of Wisconsin, Green Bay, WI 54302

In introductory group theory, the student frequently encounters examples of permutation groups G in which the elements are bijections of a set A and the group operation is composition. A few calculations show that this type of group is not (in general) commutative; that is, $f \circ g$ and $g \circ f$ need not have the same effect on the elements of A. Questions concerning the conditions under which functions do commute arise naturally, and in this note we consider the following: Are there elements of the center of G (other than the identity function e)? The existence of nontrivial elements of the center of G, or central functions, is related to the existence of functions with unique fixed points by the following result.

PROPOSITION 1. *If for each $a \in A$ there is a function $f_a \in G$ such that a is the only fixed point of f_a, then the only central function is e, the identity function.*

Proof. Let g be any central function. For $a \in A$, $g f_a = f_a g$ and, in particular, $g f_a(a) = f_a g(a)$. Since $f_a(a) = a$, we have $g(a) = f_a g(a)$, which shows that $g(a)$ is fixed by f_a. By the nature of f_a, then $g(a) = a$. Since $g(a) = a$ for all $a \in A$, we have $g = e$.

Proposition 1 may be invoked in many contexts. For example, let $A = \{1, 2, 3, \cdots, n\}$ and $G = S_n$, the group of permutations of A. The well-known result that the only central permutation is e (for $n \geqslant 3$) can be easily established by applying the proposition. It suffices to construct a permutation π_a, for each $a \in A$, that fixes a and a alone. A cyclic permutation of the elements of $A \setminus \{a\}$ evidently serves. Moreover, if $n > 3$, we may choose π_a to be an even permutation by composing with a single transposition, if necessary. Thus, the alternating group A_n of even permutations has no nontrivial central elements for $n > 3$. In a similar vein, Proposition 1 may be applied to the dihedral group, D_n, of symmetries of a regular n-gon, if n is odd. Here, A is the set of vertices, and for $a \in A$, f_a is reflection about the line through a and the center of the n-gon.

For an example with a slightly different flavor, let G consist of the differentiable bijections of the open unit interval $(0, 1)$. The function defined by $f_a(x) = (1 - x)^{r_a}$ where $r_a = (\ln a)/\ln(1 - a)$ has a as a unique fixed point. Thus, G has no central functions other than e.

In the extensive literature on permutation groups (see, for instance [3], [4], [6], or [7]), transitivity plays a major role. In a transitive group G, Proposition 1 may be applied as soon as f_a is constructed for a single $a \in A$. Then, if $b \neq a$, there is a group element g for which $g(b) = a$, and we may take $g^{-1} f_a g$ for f_b. It is clear that $g^{-1} f_a g(b) = b$. Moreover, $g^{-1} f_a g(x) = x$ leads to $f_a g(x) = g(x)$, $g(x) = a$, and finally $x = b$, so b is the unique fixed point of $g^{-1} f_a g$.

Since each example presented earlier involves a transitive group, the preceding remarks may be applied. For example, $f(x) = 1 - x$ defines a differentiable bijection on $(0, 1)$ with a unique fixed point; therefore the group of bijections on $(0, 1)$ has trivial center.

Since every permutation group G is a subdirect sum of subgroups acting transitively on the orbits of G (see [2], page 63), it is natural to seek an extension of the comments made regarding transitive groups to the general case by weakening the assumption of global transitivity. In this

regard, observe that, in the proof of Proposition 1, it is required only that f_a fix no points other than a in the orbit of a. Then, from $f_a g(a) = g(a)$ we may still conclude $g(a) = a$. Moreover, since G acts transitively on its orbits in A, if f_a exists for one element of an orbit, it exists for every element of the orbit. These remarks show that Proposition 1 generalizes to:

PROPOSITION 2. *If for each orbit $\mathcal{O} \subseteq A$ there is a function $f \in G$ with a unique fixed point in $\mathcal{O}$, then the only central function is e, the identity function.*

As an example, let G be the subgroup of S_6 generated by $(1,2)$, $(1,2,3)$, $(4,5)$, and $(4,5,6)$. Here, G has two orbits, $\{1,2,3\}$ and $\{4,5,6\}$, and is, thus, intransitive. Since $(1,2)$ fixes only 3 in the first orbit, and $(4,5)$ fixes only 6 in the second, G must have a trivial center.

Many readers will recognize that the propositions may be stated in the more general setting of a group G acting on a set A (see, for example, [**1**] or [**5**]). Informally, G acts on A if each element of G induces a bijection on A in a way that is consistent with the group product. Thus, if g and h are group elements and a is an element of A, we insist that $gh(a) = g(h(a))$ and $e(a) = a$. In this context, there is a natural homomorphism π of G into the group of permutations of A. The kernel of π may be nontrivial here, as it is in the preceding example if the full group is allowed to act on a single orbit: As far as the set $\{1,2,3\}$ is concerned, $(4,5)$ is the identity function. When the proofs of Propositions 1 and 2 are viewed in this more general setting, the conclusion states that the center of G is contained in the kernel of π. In each of the examples so far discussed π has a trivial kernel so that G has a trivial center. In our concluding example π has a nontrivial kernel and Proposition 2 is used to help characterize the center.

Let G be the group of $n \times n$ invertible matrices and A the set of one-dimensional subspaces of $\mathbb{R}^n$. If $M \in$ kernel (π), then Mx is a scalar multiple of x for each x in $\mathbb{R}^n$. Suppose $Mx = ax$, $My = by$, and $M(x+y) = c(x+y)$, where x and y are independent. Then $a = b = c$, showing that M is a scalar matrix, and that the kernel of π consists of the scalar matrices. Since scalar matrices are central in G, we also have the containment of the kernel in the center.

We now show that the center is exactly the set of scalar matrices by invoking Proposition 2. The action of G on A is transitive; A is a single orbit. Any matrix M with a single, one-dimensional eigenspace fixes a unique element of A. In particular, we may take for M the Jordan matrix with 1's on the main diagonal and superdiagonal and 0's elsewhere. Thus, Proposition 2 implies that the center is contained in the kernel, and therefore the center is the set of scalar matrices.

The author is grateful to the referee for providing insightful criticism and suggesting two of the examples.

References

1. William J. Gilbert, Modern Algebra with Applications, John Wiley, New York, 1976.
2. Marshall Hall, Jr., The Theory of Groups, Macmillan, New York, 1959.
3. B. Huppert, Endliche Gruppen, Springer-Verlag, New York, 1969.
4. Donald S. Passman, Permutation Groups, Benjamin, New York, 1968.
5. Louis W. Shapiro, Finite groups acting on sets with applications, Math. Mag., 46(1973) 136–147.
6. ______, Introduction to Abstract Algebra, McGraw-Hill, New York, 1975.
7. H. Wielandt, Finite Permutation Groups, Academic Press, New York, 1964.

DIFFERENTIAL AND DIFFERENCE EQUATIONS

RENFREY B. POTTS
Applied Mathematics Department, University of Adelaide, South Australia 5001

1. Introduction. In the vast subject of differential equations one of the elegant features is the close similarity between the theories of linear differential and linear difference equations. For example, consider the linear nonhomogeneous ODE, with constant coefficients, and of order k, written as

$$(D - m_1)(D - m_2) \cdots (D - m_k)x(t) = f(t) \tag{1.1}$$

with

$$D = \frac{dx(t)}{dt} \tag{1.2}$$

and where $m_1, m_2, \ldots, m_k$ are complex constants. Then the theorem that the general solution of (1.1) can be written as the sum of the general solution of the corresponding homogeneous equation (with $f(t) = 0$) and any particular solution $y(t)$ of (1.1) has its exact counterpart for a linear nonhomogeneous OΔE (ordinary difference equation) with constant coefficients. Such an OΔE of order k may be written

$$(E - \mu_1)(E - \mu_2) \cdots (E - \mu_k)x_n = f_n, \quad n \text{ an integer,} \tag{1.3}$$

with

$$Ex_n = x_{n+1} \tag{1.4}$$

and $\mu_1, \mu_2, \ldots, \mu_k$ complex constants.

2. Question. An important theoretical question, which seems not to have been considered, is the following:

Given the linear ODE (1.1), can a linear OΔE (1.3) be determined that has the same general solution?

By the same general solution or simply by the same solution we shall mean that for an arbitrary constant step-size h the values of x_n satisfying (1.3) are related to the solution $x(t)$ of (1.1) by

$$x_n = x(nh). \tag{2.1}$$

Geometrically this means that the x_n satisfying (1.3) will be points on the solution curve $x(t)$ of (1.1) at successive values $t = nh$, n an integer.

3. Answer. The explicit form of the general solution of (1.1) is

$$x(t) = \sum_1^k c_j \exp(m_j t) + y(t) \tag{3.1}$$

where the c_j are arbitrary constants. To try to determine a linear OΔE with the same solution, taking

$$t = nh, \; x(nh) = x_n, \; y(nh) = y_n, \tag{3.2}$$

we note immediately that since

$$E\exp(m_j nh) = \exp(m_j h)\exp(m_j nh) \tag{3.3}$$

it suffices in (1.3) to choose

$$\mu_j = \exp(m_j h) \tag{3.4}$$

and

$$f_n = (E - \mu_1)(E - \mu_2) \cdots (E - \mu_k)y_n. \tag{3.5}$$

The answer to the posed question is simply yes and the explicit form of the required OΔE is (1.3) using (3.4) and (3.5). Although we have assumed the m_j to be distinct, the conclusion is seen to be valid when some m_j coincide.

4. OΔE → ODE. It should be emphasized that the assumed constant stepsize h does not need to be "small" in any sense. Nevertheless it can be easily verified that in the limit $h \to 0$, the OΔE → the ODE. Thus from $E = \Delta + 1$ follows

$$E-\mu_j=\Delta+1-\mu_j \tag{4.1}$$

so that the OΔE (1.3) can be written

$$\prod_1^k\left(\frac{\Delta}{m_j^{-1}(\mu_j-1)}-m_j\right)x_n=\prod_1^k\left(\frac{\Delta}{m_j^{-1}(\mu_j-1)}-m_j\right)y_n. \tag{4.2}$$

Since

$$m_j^{-1}(\mu_j-1)=m_j^{-1}(e^{m_jh}-1)=h+O(h^2) \tag{4.3}$$

the OΔE (4.2) can be written, to first order in h, as

$$\prod_1^k\left(\frac{\Delta}{h}-m_j\right)x_n=\prod_1^k\left(\frac{\Delta}{h}-m_j\right)y_n, \tag{4.4}$$

giving in the limit $h\to 0$

$$\prod_1^k(D-m_j)x(t)=\prod_1^k(D-m_j)y(t)=f(t). \tag{4.5}$$

The special cases of equal m's or zero m's cause no difficulties.

5. Examples. The significance of the question and answer above is best illustrated by elementary examples of ODE's and OΔE's with the same solutions.

(1)
$$\text{ODE}\quad \dot x-mx=0,\quad m\neq 0, \tag{5.1}$$

$$\text{OΔE}\quad (E-e^{mh})x_n=0, \tag{5.2}$$

or
$$\frac{x_{n+1}-x_n}{m^{-1}(e^{mh}-1)}-mx_n=0. \tag{5.3}$$

The general solution of (5.1) is

$$x(t)=e^{mt}c$$

with c an arbitrary constant while the general solution of (5.2) or (5.3) is

$$x_n=e^{mnh}c. \tag{5.4}$$

It is in the sense that $x_n=x(nh)$ that the ODE and the OΔE are described as having the same general solution. This is true regardless of the sign or magnitude of h.

The usual OΔE approximation to (5.1), namely,

$$\frac{x_{n+1}-x_n}{h}-mx_n=0, \tag{5.5}$$

does not enjoy this property.

(2)
$$\text{ODE}\quad \ddot x+\omega^2x=\sin\gamma t,\qquad \gamma\neq\omega, \tag{5.6}$$

$$\text{OΔE}\quad (E^2-2\cos\omega hE+1)x_{n-1}=(E^2-2\cos\omega hE+1)\frac{\sin\gamma(n-1)h}{\omega^2-\gamma^2}, \tag{5.7}$$

or
$$\frac{x_{n+1}-2x_n+x_{n-1}}{4\omega^{-2}\sin^2\frac12\omega h}+\omega^2x_n=\frac{\omega^2}{\omega^2-\gamma^2}\left[1-\frac{\sin^2\frac12\gamma h}{\sin^2\frac12\omega h}\right]\sin\gamma nh. \tag{5.8}$$

The general solution of (5.6) is

$$x(t)=c_1\cos\omega t+c_2\sin\omega t+\frac{1}{\omega^2-\gamma^2}\sin\gamma t \tag{5.9}$$

while that of (5.7) or (5.8) is

$$x_n = c_1\cos\omega nh + c_2\sin\omega nh + \frac{1}{\omega^2-\gamma^2}\sin\gamma nh \tag{5.10}$$

so that again $x_n = x(nh)$, regardless of the magnitude of h. Note that it is convenient to use x_{n-1} in (5.7) rather than x_n and that the denominator $4\omega^{-2}\sin^2\frac{1}{2}\omega h$ in the first term of (5.8) replaces the usual h^2; this replacement is essential to make the general solutions of the ODE and the OΔE the same.

(3) ODE $\quad \ddot{x} + \omega^2 x = \sin\omega t, \qquad (5.11)$

$$\text{O}\Delta\text{E} \quad (E^2 - 2\cos\omega hE + 1)x_{n-1} = (E^2 - 2\cos\omega hE + 1)\frac{-h(n-1)}{2\omega}\cos\omega(n-1)h, \tag{5.12}$$

$$\text{or} \quad \frac{x_{n+1} - 2x_n + x_{n-1}}{4\omega^{-2}\sin^2\frac{1}{2}\omega h} + \omega^2 x_n = \frac{\omega h}{2\tan\frac{1}{2}\omega h}\sin\omega nh. \tag{5.13}$$

The general solution of (5.11) is

$$x(t) = c_1\cos\omega t + c_2\sin\omega t - \frac{t}{2\omega}\cos\omega t \tag{5.14}$$

while that of (5.12) or (5.13) is

$$x_n = c_1\cos\omega nh + c_2\sin\omega nh - \frac{nh}{2\omega}\cos\omega nh, \tag{5.15}$$

so that again

$$x_n = x(nh).$$

For further examples it is left to the reader to take his favorite ODE and use (1.3) with (3.4) and (3.5) to find the related OΔE.

6. System of equations. The same approach can be used to determine the system of linear OΔE's with the same solution as the system of linear ODE's

$$Dx(t) = Mx(t), \tag{6.1}$$

where M is a constant $r \times r$ matrix. The explicit solution of (6.1) is

$$x(t) = \exp(Mt)x(0). \tag{6.2}$$

The correct system of OΔE's to choose is

$$Ex_n = \exp(Mh)x_n \tag{6.3}$$

with solution

$$x_n = \exp(Mnh)x_0. \tag{6.4}$$

Again this is valid for any constant step size h.

For small h we can first rewrite (6.3) as

$$\Delta x_n = (\exp(Mh) - I)x_n, \tag{6.5}$$

which to first order gives

$$\frac{\Delta x_n}{h} = Mx_n, \tag{6.6}$$

an approximation to (6.1), which is the one usually taken.

The results in the theory of stability of the system (6.1) are directly applicable to (6.3) but not to (6.6).

7. Comments. The simple result presented above relating linear differential and linear difference equations requires some comments.

(1) In making the particular choice of ΔE from the infinite number of possible ΔE's that approximate a given DE, the approach has been different from that adopted in numerical analysis where emphasis is given to high-order approximations, stability, and control of round-off errors. The aim has been to choose a ΔE that has the same solution as the DE and it is assumed that the ΔE is exactly solvable, regardless of step size h.

(2) The DE's and ΔE's exhibited have the same general solutions involving arbitrary constants c_j. Boundary conditions for a DE that prescribe values of $x(t)$ at points that can be chosen as exact grid points are immediately applicable to the corresponding ΔE, but conditions prescribing values of the derivatives of $x(t)$ require some manipulation. For example, the initial-value problem

$$(D-m_1)(D-m_2)x(t)=0, \qquad m_1 \neq m_2, \tag{7.1}$$

given

$$x(0), \qquad \dot{x}(0) \tag{7.2}$$

has the same solution as

$$(E-e^{m_1 h})(E-e^{m_2 h})x_n=0 \tag{7.3}$$

with

$$\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ e^{m_1 h} & e^{m_2 h} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ m_1 & m_2 \end{bmatrix}^{-1} \begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix}, \tag{7.4}$$

or, explicitly,

$$x_0 = x(0)$$

and

$$x_1 = \frac{m_2 e^{m_1 h} - m_1 e^{m_2 h}}{m_2 - m_1} x(0) + \frac{e^{m_2 h} - e^{m_1 h}}{m_2 - m_1} \dot{x}(0). \tag{7.5}$$

For small h, note that (7.5) gives

$$x_1 = x_0 + h\dot{x}(0) + 0(h^2) \tag{7.6}$$

as expected. The generalization to equations of higher order is immediate.

(3) It is clear that the approach used in this paper can be extended to some linear ODE's with nonconstant coefficients and to some PDE's. But much current interest is centered on nonlinear equations [**1**], [**2**]. The interesting review article by May [**2**] details some of the fascinating and surprising behavior of simple first-order nonlinear difference equations and in recent papers [**3**], [**4**] the present author has shown how the classical nonlinear problems described by the Duffing, Verhulst and Volterra differential equations can be analyzed by difference equations. An important application of the present work on linear equations is to problems in which the nonlinear effects can be considered as perturbations. The results obtained show how the unperturbed linear DE might be best replaced by a linear ΔE to which the nonlinear perturbations could then be added. For example, in choosing a ΔE to approximate Van der Pol's DE

$$\ddot{x} - \varepsilon(1-x^2)\dot{x} + x = 0 \tag{7.7}$$

the replacement

$$\ddot{x} \rightarrow \frac{x_{n+1} - 2x_n + x_{n-1}}{4\sin^2\frac{1}{2}h} \tag{7.8}$$

forces the unperturbed period to have the correct value 2π. The usual replacement with h^2 in the denominator of (7.8) does not achieve this.

Acknowledgements. The author has greatly benefitted from discussions with Professor Elliott Montroll, and he has enjoyed the hospitality of Professor Joe Keller and his colleagues in the Department of Mathematics at Stanford University where the reported work has been carried out.

References

1. R. Hirota, Nonlinear partial difference equations, V. Nonlinear equations reducible to linear equations, J. Phys. Soc. Japan, 46 (1979) 312–319.
2. R. M. May, Simple mathematical models with very complicated dynamics, Nature, 261 (1976) 459–467.
3. R. B. Potts, Exact solution of a difference approximation to Duffing's equation, J. Aust. Math. Soc. B, 23 (1980) 64–77.
4. ______, Nonlinear difference equations, J. Nonlin. An. (in press).

OUTER AUTOMORPHISMS OF S_6

GERALD JANUSZ AND JOSEPH ROTMAN
Department of Mathematics, University of Illinois, Urbana, IL 61801

For a set X, let S_X denote the group of all permutations of X; when $X = \{1,2,\ldots,n\}$, write S_n instead of S_X. Each element a of a group G gives an automorphism γ_a of G (called *conjugation* by a) defined by $\gamma_a(g) = aga^{-1}$; an automorphism of this form is called *inner*; otherwise an automorphism is called *outer*. The function $G \rightarrow \mathrm{Aut}(G)$ defined by $a \mapsto \gamma_a$ is a homomorphism with kernel $Z(G)$, the center of G, and image $\mathrm{Inn}(G)$, the subgroup consisting of all inner automorphisms. If $Z(G) = \{1\}$ and $\mathrm{Inn}(G) = \mathrm{Aut}(G)$, then G is called *complete*. It follows from our remarks that, when G is complete, $G \cong \mathrm{Aut}(G)$. It is well known [**4**; Theorem 7.4] that S_n is complete for $n \neq 2$ and $n \neq 6$. As S_2 has order two and everything is known, let us examine S_6. It is an easy exercise to show $Z(S_n) = \{1\}$ for $n \geqslant 3$. Hölder [**2**] first proved (in 1895) that S_6 has an outer automorphism that is unique up to multiplication by an inner automorphism; he showed $\mathrm{Aut}(S_6)/\mathrm{Inn}(S_6) \cong \mathbf{Z}/2\mathbf{Z}$. We offer a constructive proof of this fact and a bit more; we show that $\mathrm{Aut}(S_6)$ is a semidirect product of $\mathrm{Inn}(S_6)$ ($\cong S_6$) by a group of order two.

The main tool we use, apart from standard elementary results about symmetric groups, is the following:

If P is a subgroup of G and if X is the family of conjugates of P in G, then there is a homomorphism $\rho : G \rightarrow S_X$ defined, for $a \in G$, by

$$\rho(a) : gPg^{-1} \mapsto agPg^{-1}a^{-1}.$$

This homomorphism has kernel $\cap_{a \in G} aN(P)a^{-1}$ where $N(P)$ is the normalizer of P in G. Call ρ *the representation of G on the conjugates of P.*

Here is an easy illustration of the use of this tool.

LEMMA 1. *A subgroup K of S_6 having order* 120 *is equal to its normalizer*: $K = N(K)$; *moreover K has exactly six conjugates.*

Proof. If K has k conjugates, then $k \leqslant 6$ because

$$k = [S_6 : N(K)] \leqslant [S_6 : K] = 6!/120 = 6.$$

If $k = 6$ we are done. Suppose $k < 6$. The representation of S_6 on the conjugates of K gives a homomorphism $\rho : S_6 \rightarrow S_k$ and $\ker \rho \subseteq N(K)$. The kernel of ρ is a normal subgroup and so equals $\{1\}$, A_6, or S_6, as these are the only normal subgroups. If $\ker \rho = \{1\}$, then

S_k has a subgroup (image ρ) with 6! elements; this is impossible when $k < 6$. If $\ker \rho = A_6$, then $N(K) = A_6$, and so K is a normal subgroup of A_6, contradicting the simplicity of A_6. If $\ker \rho = S_6$, then K is normal in S_6, which is also impossible since no normal subgroup has order 120. □

A subgroup K of S_X is *transitive* if, for every pair of elements $x, y \in X$, there is a permutation $\sigma \in K$ with $\sigma(x) = y$. If $\rho : G \to S_X$ is the representation of G on the conjugates of a subgroup P of G, then $\operatorname{im} \rho$ is easily seen to be a transitive subgroup of S_X.

LEMMA 2. *If K is a transitive subgroup of S_6 having order* 120, *then K cannot contain a transposition.*

Proof. First of all, K contains an element α of order 5, which must be a 5-cycle, say, $\alpha = (12345)$. If $(ij) \in K$, then transitivity of K provides $\beta \in K$ with $\beta(j) = 6$. Therefore $\beta(ij)\beta^{-1} = (l6)$ for some $l(l \neq 6)$. Conjugating $(l6)$ by the powers of α shows K contains $(16), (26), (36), (46), (56)$. But it is a standard exercise that these transpositions generate all of S_6. □

THEOREM 3. *There exists an outer automorphism of S_6.*

Proof. We begin by showing S_6 contains a transitive subgroup K of order 120. Now a Sylow 5-subgroup P of S_5 must have 6 conjugates (by the Sylow theorems). The representation ρ of S_5 on the conjugates of P gives a homomorphism $\rho : S_5 \to S_6$, which must be one-one, for $\ker \rho \subseteq N(P)$ and hence is not one of the subgroups A_5 or S_5. Hence $K = \operatorname{im} \rho$ is a transitive subgroup of S_6 having order 120 and index 6.

Now let $\phi : S_6 \to S_6$ be the representation on the conjugates of K. The same argument just given shows ϕ is one-one, hence onto and so $\phi \in \operatorname{Aut}(S_6)$. If ϕ were an inner automorphism, then $\phi((12))$ would be a transposition (thus fixing four symbols) and so (12) would normalize exactly four conjugates of K. If $\alpha \in S_6$ and $(12)\alpha K \alpha^{-1}(12) = \alpha K \alpha^{-1}$, then $\alpha^{-1}(12)\alpha$ lies in $N(K)$. By Lemma 1, $K = N(K)$ and so K contains the transposition $\alpha^{-1}(12)\alpha$, which contradicts Lemma 2. Thus ϕ is an outer automorphism. □

The proof of the essential uniqueness of ϕ is facilitated by the following table describing S_6. Recall that two permutations lie in the same conjugacy class if and only if they have the same cycle structure.

	cycle structure	order	parity	number of such
C_1	(1)	1	even	1
C_2	(12)	2	odd	15
C_3	(123)	3	even	40
C_4	(1234)	4	odd	90
C_5	(12345)	5	even	144
C_6	(123456)	6	odd	120
C_7	(13)(34)	2	even	45
C_8	(12)(345)	6	odd	120
C_9	(12)(3456)	4	even	90
C_{10}	(12)(34)(56)	2	odd	15
C_{11}	(123)(456)	3	even	40
				$720 = 6!$

THEOREM 4. $\operatorname{Aut}(S_6)/\operatorname{Inn}(S_6) \cong \mathbf{Z}/2\mathbf{Z}$.

Proof. If $Y = \{C_1, C_2, \ldots, C_{11}\}$ is the set of conjugacy classes of S_6 and if $\phi \in \operatorname{Aut}(S_6)$, then $\phi \in S_Y$. Now ϕ is inner if and only if $\phi(C_2) = C_2$ [**4**; Lemma 7.3]; therefore, ϕ is outer if and only if ϕ interchanges C_2 and C_{10}, these being the only conjugacy classes having 15 elements. It follows

that if ϕ and ψ are outer automorphisms, then $\phi\psi(C_2) = C_2$, whence $\phi\psi$ is inner, and $\text{Aut}(S_6)/\text{Inn}(S_6)$ has order 2. □

Now we turn to an explicit construction of an outer automorphism ϕ. We know of four other such descriptions. In [**1**] Bender notes that S_6 has a presentation in terms of specific generators and relations. He produces two sets of generators that satisfy the relations and thereby obtains an outer automorphism. In [**3**], Miller defines ϕ by directly defining $\phi(1i)$ for $2 \leqslant i \leqslant 6$ and showing this determines a well-defined outer automorphism. The other two proofs make use of unusual isomorphic copies of S_6. In [**5**], Witt proves the Mathieu group M_{12} contains a subgroup T isomorphic to S_6 and an element σ that normalizes T and induces an outer automorphism of T. Finally, the group $P\Gamma L(2,9)$, the group of all nonsingular semilinear fractional transformations over $GF(9)$, contains a subgroup U isomorphic to S_6 and an element τ that normalizes U and induces an outer automorphism of U. (This last fact was pointed out to us by J. Walter.)

In contrast to the methods just mentioned, we present an explicit construction of an outer automorphism ϕ. It s clear that if $X = \{h_1,\dots,h_n\}$, then the obvious one-one correspondence $f\colon X \to \{1,2,\dots,n\}$ given by $f(h_i) = i$ induces an isomorphism $S_X \to S_n$ by $\alpha \mapsto f\circ\alpha\circ f^{-1}$. In words, *relabeling* the set X produces an isomorphism. The idea now is to find convenient relabelings of the various sets of six elements that have arisen.

First of all, we view S_6 as permutations of $\{1,2,\dots,6\}$ and S_5 as its subgroup consisting of all permutations of $\{1,\dots,5\}$, i.e., those permutations that fix 6. Let $\sigma = (12345)$ and let P_6 be the subgroup of S_5 generated by σ. Define

$$P_1 = (45)P_6(45)$$

and, for $1 \leqslant j \leqslant 4$,

$$P_{j+1} = \sigma^j P_1 \sigma^{-j}.$$

Thus, conjugation by σ cyclically permutes $P_1,\dots,P_5$ and fixes P_6. If $X = \{P_1,\dots,P_6\}$ and if $\alpha \in S_5$, our construction in Theorem 3 identifies S_X with S_6 by relabeling P_i as i; that is,

$$\rho(\alpha) = \begin{pmatrix} P_i \\ \alpha P_i \alpha^{-1} \end{pmatrix} = \begin{pmatrix} i \\ j \end{pmatrix}$$

where $\alpha P_i \alpha^{-1} = P_j$. Under this identification, our remarks above identify $\rho(\sigma)$ with σ. Indeed, for $\alpha \in S_5$, $\rho(\alpha) \in K \cap S_5$ (where $K = \text{Im}\,\rho$) if and only if $\rho(\alpha)$ leaves 6 fixed, i.e., $\alpha P_6 \alpha^{-1} = P_6$, whence $\alpha \in N$, the normalizer of P_6 in S_5. Therefore $K \cap S_5 = \rho(N)$. The subgroup N has a unique Sylow 5-subgroup, namely P_6, since any subgroup is normal in its normalizer. It follows that $\rho(N)$ has just one Sylow 5-subgroup, namely, $\rho(P_6) = P_6$. Label the six conjugates of K in S_6 in the same fashion that we labeled the conjugates of P_6:

$$K_6 = K; \quad K_1 = (45)K(45); \quad K_{j+1} = \sigma^j K_1 \sigma^{-j} \text{ for } \quad 1 \leqslant j \leqslant 4.$$

These are distinct conjugates because $P_i \subset K_i$ for all i, and no K_i can contain more than one of the P_j.

Here, following, is the main computational device that enables us to give a formula for the outer automorphism ϕ of Theorem 3.

LEMMA 5. *If $\alpha \in S_5$, then $\alpha P_i \alpha^{-1} = P_j$ if and only if $\alpha K_i \alpha^{-1} = K_j$. In other words, with the obvious relabeling as permutations of $\{1,2,\dots,6\}$,*

$$\rho(\alpha) = \phi(\alpha).$$

Proof. If $\alpha P_i \alpha^{-1} = P_j$, then $P_j \subset \alpha K_i \alpha^{-1}$. But we have chosen the labels so that K_j is the conjugate containing P_j. □

The main point of Lemma 6 is that one may painlessly compute $\phi(\alpha)$ for $\alpha \in S_5$, and it is

unnecessary to determine the conjugates K_i completely. Unfortunately, determination of $\phi(\alpha)$ for α not in S_5 involves some calculations.

LEMMA 6. *The outer automorphism* ϕ *has the following values*:

$$\phi((12)) = (12)(36)(45)$$
$$\phi((13)) = (16)(24)(35)$$
$$\phi((14)) = (13)(25)(46)$$
$$\phi((15)) = (15)(26)(34)$$
$$\phi((16)) = (14)(23)(56)$$

Proof. The value of $\phi((1i))$ for $2 \leqslant i \leqslant 5$ is found, using Lemma 6, by determining $(1i)P_j(1i)$ for each j. This short computation is left to the reader.

The determination of $\phi((16))$ may be carried out as follows: Make a list of generators $\sigma_1, \dots, \sigma_6$ of the six Sylow 5-subgroups of S_5. Then $\rho(\sigma_1), \dots, \rho(\sigma_6)$ generate the six Sylow 5-subgroups of K_6. Next, conjugate these by (45) and by the powers of σ to obtain generators of the Sylow 5-subgroups in each K_i. Finally, select some K_i, some generator β of one of its Sylow 5-subgroups, form $(16)\beta(16)$, and locate the (unique) K_j in which it lies. Then $\phi((16))$ carries i to j. The result of this work is stated in the lemma. □

Since every element of S_6 is a product of transpositions of the form $(1i)$, the information above allows one to evaluate $\phi(\beta)$ for every $\beta \in S_6$.

COROLLARY 7. *If* ϕ *is an outer automorphism of* S_6, *then* ϕ *takes a* 3-*cycle to a product of two disjoint* 3-*cycles, but* ϕ *preserves the cycle structure of any permutation of order* 4.

Proof. Just evaluate ϕ on $(123) = (13)(12)$ and on $(1234) = (14)(13)(12)$. □

COROLLARY 8. *If* ϕ *is the automorphism of Theorem* 3, *then*

$$\phi^2 = \gamma_{\sigma^{-2}}, \quad \text{conjugation by} \quad \sigma^{-2}.$$

Proof. Using Lemma 7, one may show that $\phi^2(\beta) = \sigma^{-2}\beta\sigma^2$ whenever $\beta = (1i)$ and $2 \leqslant i \leqslant 6$; it follows that this formula holds for every $\beta \in S_6$. □

THEOREM 9. $\mathrm{Aut}(S_6) = \mathrm{Inn}(S_6)\langle\psi\rangle$ *with* ψ *an outer automorphism of order* 2. *Therefore,* $\mathrm{Aut}(S_6)$ *is a group of order* 1440 *that is a semidirect product of* S_6 *by a group of order* 2.

Proof. Define $\psi = \gamma_\sigma\phi$. It suffices to show ψ has order 2. If $\beta \in S_6$, then

$$\psi^2(\beta) = \gamma_\sigma\phi\gamma_\sigma\phi(\beta) = \gamma_\sigma\phi\big(\sigma\phi(\beta)\sigma^{-1}\big) = \gamma_\sigma\big(\phi(\sigma)\phi^2(\beta)\phi(\sigma^{-1})\big).$$

Since $\phi(\sigma) = \sigma$ (as we saw before Lemma 6), Corollary 8 shows these equations conclude with $\psi^2(\beta) = \beta$. □

References

1. H. A. Bender, A new method for the determination of the group of isomorphisms of the symmetric group of degree n, this MONTHLY, 31(1924) 287–289.

2. O. Hölder, Bildung zusammengesetzter Gruppen, Math. Ann., 46(1895) 321–422.

3. D. W. Miller, On a theorem of Hölder, this MONTHLY, 65(1958) 252–254.

4. J. J. Rotman, The Theory of Groups, An Introduction, 2nd ed., Allyn & Bacon, Boston, MA, 1973.

5. E. Witt, Die 5-fach transitiven Gruppen von Mathieu, Abh. Math. Sem. Univ. Hamburg, 12(1938) 256–264.

ANSWER TO "PHOTO" ON PAGE 376

Leonidas Alaoglu.

A NATURALLY OCCURRING FUNCTION CONTINUOUS ONLY AT IRRATIONALS

AMITAVA BAGCHI
Indian Institute of Management Calcutta, P.O. Box 16757, Calcutta-700027, India

EDWARD M. REINGOLD
Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801

In this note we present a function f: $D \to R$ with the properties that
(i) the domain $D = [0, 1/2]$,
(ii) the range R is a subset of $[0, 1]$,
(iii) f is strictly decreasing, and
(iv) for any $\alpha \in D$, f is continuous at α if α is irrational, and f is discontinuous at α if α is rational.

Such peculiarities are well known in real analysis (see, for example, [**1**]), but what makes f remarkable is that it occurs quite naturally in the analysis of an important data structure from computer science.

The function f arises in connection with random binary search trees. We shall only sketch the basic definitions and relevant background material, referring the reader to [**2**] or [**3**] for a comprehensive presentation.

A *binary tree* is a finite set of *nodes* that is either *empty* or consists of a *root* and two disjoint binary trees called the *left* and *right subtrees* of the root. It is customary to draw such a tree as shown in Fig. 1.

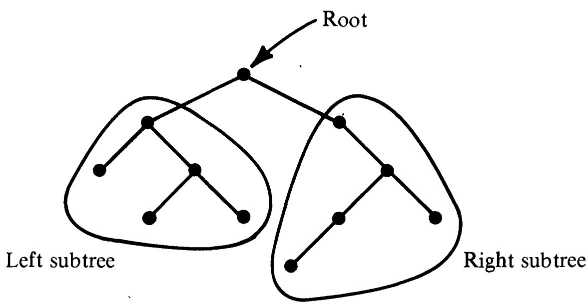


FIG. 1. A binary tree.

A sequence of distinct numbers

$$S = \{s_1, s_2, \ldots, s_n\}, \quad n \geq 0,$$

defines a binary tree $T(S)$ as follows: If $n = 0$, $T(S)$ is the empty tree; if $n \geq 1$, $T(S)$ consists of the root node s_1, a left subtree $T(S_L)$, and a right subtree $T(S_R)$, where

$$S_L = \{s_{i_1}, s_{i_2}, \ldots, s_{i_l}\}, \quad i_1 < i_2 < \cdots < i_l,$$

is the subsequence of S of all elements smaller than s_1 and

$$S_R = \{s_{j_1}, s_{j_2}, \ldots, s_{j_r}\}, \quad j_1 < j_2 < \cdots < j_r,$$

is the subsequence of S of all elements larger than s_1. The tree of Fig. 1, for example, is defined by the sequence

$$S = \{6, 2, 4, 7, 10, 11, 1, 3, 9, 5, 8\}$$

for which

$$S_L = \{2, 4, 1, 3, 5\}$$

and

$$S_R = \{7, 10, 11, 9, 8\}.$$

The sequence defining a particular tree is *not* unique; thus

$$S' = \{6, 7, 10, 11, 2, 1, 4, 5, 3, 9, 8\},$$

also, defines the tree of Fig. 1, as do many other sequences.

The *balance* of a node r in a binary tree is defined as the ratio

$$\frac{1 + \text{number of nodes in the left subtree of } r}{1 + \text{number of nodes in the subtree whose root is } r}.$$

In Fig. 2 we show the binary tree of Fig. 1 with each node labeled with its balance. The balances describe the "shape" of the tree—a node with balance close to $1/2$ means that the subtree rooted at that node is relatively balanced, while a balance close to 0 or 1 means the subtree is relatively unbalanced.

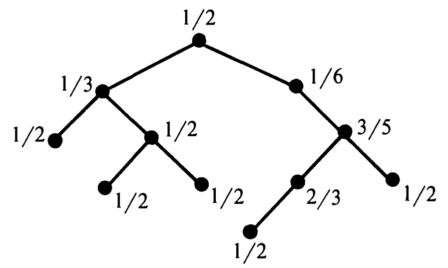


FIG. 2. The binary tree of Fig. 1 labeled with balances.

Binary trees are an important data structure in computer science for organizing information for efficient retrieval. The extent of the balance or imbalance in a binary tree determines the efficiency of the retrieval process. Essentially, "balanced" binary trees have *height* (length of the longest path from the root) that is logarithmic in the number of nodes, while "unbalanced" trees can have height linear in the number of nodes. The height, it turns out, is an excellent measure of the retrieval time.

It is therefore of interest to determine the expected shape of binary trees constructed at random; we express this by asking: What is the expected distribution of balances in a random binary tree? That is, given any real $\alpha \in [0, 1/2]$, what is the expected fraction of nodes with balance in the range $[\alpha, 1-\alpha]$? It is this question that leads to a function with the properties described above.

By a *random binary tree* we mean the binary tree $T(S)$ defined by a random permutation of $S = \{1, 2, \ldots, n\}$ where each of the $n!$ permutations of S is equally probable. The nonuniqueness of the defining sequences causes a nonuniform distribution over binary trees. Although this distribution may seem strange taken out of context, it is much more meaningful in terms of the information retrieval application than the uniform distribution over binary trees. In that application, we want to consider a uniform random distribution over possible "input" sequences, not tree shapes.

Let T be a random binary tree of n nodes and let $\alpha, 0 < \alpha \leqslant 1/2$, be given. Let a_n be the expected number of nodes in T with balances in the range $[\alpha, 1-\alpha]$. We have $a_0 = 0$ and, since all $n!$ permutations of $\{1, 2, \ldots, n\}$ are equally probable, for $n \geqslant 1$ we have

$$na_n = \sum_{i=1}^{n} (a_{i-1} + a_{n-i}) + \lfloor (1-\alpha)(n+1) \rfloor - \lceil \alpha(n+1) \rceil + 1, \tag{1}$$

where $\lfloor x \rfloor$, read "floor of x," is the greatest integer less than or equal to x, and $\lceil x \rceil$, read "ceiling

of x," is the least integer greater than or equal to x. We get (1) by observing that the balance of the root node of T is in the required range if

$$\lfloor(1-\alpha)(n+1)\rfloor \geqslant s_1 \geqslant \lceil\alpha(n+1)\rceil,$$

and that the probability of any particular s_1 is $1/n$. From (1), for $n \geqslant 1$, we have

$$\begin{aligned}(n+1)a_{n+1}-(n+2)a_n = &\lfloor(n+2)(1-\alpha)\rfloor - \lceil(n+2)\alpha\rceil \\ &-\lfloor(n+1)(1-\alpha)\rfloor + \lceil(n+1)\alpha\rceil. \end{aligned} \tag{2}$$

Define $f(\alpha)$ to be the limiting value of the expected fraction of the nodes with balance in the range $[\alpha, 1-\alpha]$ in a random binary tree, that is,

$$f(\alpha) = \lim_{n\to\infty} \frac{a_n}{n}. \tag{3}$$

This limit exists by the Cauchy criterion since (2) gives

$$\frac{a_{n+1}}{n+2} - \frac{a_n}{n+1} = \frac{\varepsilon_n}{(n+1)(n+2)}$$

where $\varepsilon_n = \pm 1$, so that, for $m \geqslant n$,

$$\frac{a_m}{m+1} - \frac{a_n}{n+1} \leqslant \frac{1}{n+1}.$$

Furthermore, this proves that the convergence is uniform for $0 < \alpha \leqslant 1/2$.

We get the desired distribution of balances by solving (2) and (3) for rational values of $\alpha > 0$. Clearly, $f(0) = 1$ since *all* nodes have balance in the range [0, 1]. Let

$$\alpha = p/q, \quad 0 < \alpha \leqslant 1/2,$$

where p and q are positive integers, not necessarily relatively prime. To solve (2), we define the generating function

$$A(z) = \sum_{n=0}^{\infty} a_n z^n.$$

In (2), the right-hand side is -1 if

$$n+1 = \lfloor kq/p \rfloor$$

where k is any positive integer, and the right-hand side is $+1$ otherwise. So, multiplying each side of (2) by z^n and summing gives

$$\sum_{n=1}^{\infty}(n+1)a_{n+1}z^n - \sum_{n=1}^{\infty}(n+2)a_n z^n = \sum_{n=1}^{\infty} z^n - 2\sum_{k=1}^{\infty} z^{\lfloor kq/p\rfloor - 1},$$

which becomes the differential equation

$$A' - \frac{2}{1-z}A = \frac{1}{(1-z)^2} - 2\sum_{k=1}^{p} \frac{z^{j_k-1}}{(1-z^q)(1-z)}, \tag{4}$$

where

$$j_k = \lfloor kq/p \rfloor, \quad 1 \leqslant k \leqslant p.$$

To determine a_n, we need a series solution of (4) in powers of z. The right-hand side of (4) has a Maclaurin series expansion which is valid for $|z| < 1$, so (4) has a series solution in powers of z valid for $|z| < 1$ (see, e.g., [4]). We can rewrite (4) in the form

$$\frac{d}{dz}\left[(1-z)^2 A\right] = 1 - 2\sum_{k=1}^{p} \frac{(1-z)z^{j_k-1}}{1-z^q}$$

$$= 1 - 2 \sum_{k=1}^{p} \sum_{i=0}^{\infty} \left(z^{iq+j_k-1} - z^{iq+j_k} \right).$$

Integrating, dividing by $(1-z)^2$, and using the boundary condition $a_0 = 0$ [i.e., $A(0) = 0$], we get

$$A(z) = \frac{z}{(1-z)^2} - \frac{2}{(1-z)^2} \sum_{k=1}^{p} \sum_{i=0}^{\infty} \left(\frac{z^{iq+j_k}}{iq+j_k} - \frac{z^{iq+j_k+1}}{iq+j_k+1} \right).$$

Express n as $n = i_0 q + j_{k_0}$. Then

$$a_n = n - 2(n+1) \sum_{k=1}^{p} \sum_{i=0}^{i_0-1} \frac{1}{(iq+j_k)(iq+j_k+1)} - 2(n+1) \sum_{k=1}^{k_0} \frac{1}{(i_0 q + j_k)(i_0 q + j_k + 1)}.$$

Combining this with (3) we have

$$f(\alpha) = 1 - 2 \sum_{k=1}^{p} \sum_{i=0}^{\infty} \frac{1}{(iq+j_k)(iq+j_k+1)}, \quad 0 < \alpha \leqslant 1/2. \tag{5}$$

To determine $f(\alpha)$ for irrational α, we observe that no node in a binary tree can ever have an irrational balance. Let

$$0 < r_1 < r_2 < r_3 < \cdots$$

be an infinite sequence of rationals such that

$$\lim_{i \to \infty} r_i = \alpha.$$

Define

$$f^-(\alpha) = \lim_{i \to \infty} f(r_i). \tag{6}$$

Because a_n as a function of α is monotonic nonincreasing, f must be monotonic nonincreasing in α. Hence $f(r_1), f(r_2), f(r_3), \cdots$ is a bounded, nonincreasing sequence of real numbers, so that the limit in (6) exists. Furthermore, the limit is the same for *any* such sequence of rationals: Given any two such sequences, "merging" them gives a supersequence of each; the limit (6) exists for the supersequence as above. Since each of the given sequences is a subsequence of the supersequence, it follows that their limits are, in each case, the same as the limit for the supersequence; hence they are equal. $f^-(\alpha)$ is thus well defined, irrespective of whether α is rational or irrational.

We now observe that for all $\alpha, 0 < \alpha < 1/2$,

$$f(\alpha) = f^-(\alpha).$$

This follows from

$$\begin{aligned} f(\alpha) - f^-(\alpha) &= f(\alpha) - \lim_{i \to \infty} \lim_{n \to \infty} \frac{a_n(r_i)}{n} \\ &= \lim_{i \to \infty} \left[f(\alpha) - \lim_{n \to \infty} \frac{a_n(r_i)}{n} \right] \\ &= \lim_{i \to \infty} \lim_{n \to \infty} \frac{a_n(\alpha) - a_n(r_i)}{n} \\ &= \lim_{n \to \infty} \frac{1}{n} \lim_{i \to \infty} \left[a_n(\alpha) - a_n(r_i) \right], \end{aligned}$$

the interchange of limits being justified by the uniform convergence of a_n/n. The inner limit is zero because, for $\alpha - r_i$ sufficiently small, that is, for i sufficiently large, $a_n(\alpha) = a_n(r_i)$ since there will be no nodes with balance in $[r_i, \alpha)$ or $(1-\alpha, 1-r_i]$ in a tree of n nodes.

Let

$$1/2 > s_1 > s_2 > s_3 \cdots$$

be an infinite sequence of rationals such that

$$\lim_{i\to\infty} s_i = \alpha.$$

Define

$$f^+(\alpha) = \lim_{i\to\infty} f(s_i).$$

As with $f^-(\alpha)$, $f^+(\alpha)$ is well defined.

We are now ready to establish that f has the properties stated at the beginning of this paper. Properties (i) and (ii) are clearly satisfied. We defer the proof of property (iii) until after establishing property (iv).

For fixed α, let b_n be the expected number of nodes in an n-node random binary tree with balance *exactly* α or $1-\alpha$. If α is irrational, clearly $b_n = 0$ for all n. For rational $\alpha = p/q < 1/2$ with *p and q relatively prime*, we have, in a manner similar to the derivation of a_n,

$$nb_n = \begin{cases} \sum_{i=1}^{n} (b_{i-1} + b_{n-i}) + 2 & \text{if } n \equiv 0 \bmod q, \\ \sum_{i=1}^{n} (b_{i-1} + b_{n-i}) & \text{otherwise,} \end{cases} \tag{7}$$

for $n \geqslant 1$, and $b_0 = 0$. Define $\Delta(\alpha)$ to be the limiting value of the expected fraction of the nodes in a random binary tree with balance exactly α or $1-\alpha$, that is,

$$\Delta(\alpha) = \lim_{n\to\infty} \frac{b_n}{n}.$$

As in the manipulations of a_n and f, we consider the difference $(n+1)b_{n+1} - nb_n$ to transform (7) into a differential equation on the associated generating function: For $n \geqslant 1$ we have

$$(n+1)b_{n+1} - (n+2)b_n = \begin{cases} 2 & \text{if } n+2 \equiv 0 \bmod q, \\ -2 & \text{if } n+1 \equiv 0 \bmod q, \\ 0 & \text{otherwise.} \end{cases}$$

Multiplying each side by z^n and summing gives

$$\sum_{n=1}^{\infty} (n+1)b_{n+1}z^n - \sum_{n=1}^{\infty} (n+2)b_n z^n = 2(1-z)\sum_{k=1}^{\infty} z^{kq-2},$$

which yields the differential equation

$$B' - \frac{2}{1-z}B = \frac{2z^{q-2}}{1-z^q},$$

where $B(z) = \sum_{n=0}^{\infty} b_n z^n$. This differential equation has, also, a series solution in powers of z valid for $|z| < 1$. Solving as with $A(z)$, we find

$$B(z) = \frac{2}{(1-z)^2}\left[\sum_{i=1}^{\infty} \frac{z^{iq-1}}{iq-1} - 2\sum_{i=1}^{\infty} \frac{z^{iq}}{iq} + \sum_{i=1}^{\infty} \frac{z^{iq+1}}{iq+1}\right],$$

so that

$$\Delta(\alpha) = 4 \sum_{i=1}^{\infty} \frac{1}{(iq-1)iq(iq+1)}. \tag{8}$$

From the definitions of $f^-(\alpha)$ and $f^+(\alpha)$ and the uniform convergence of a_n/n, it follows that $f^-(\alpha) - f^+(\alpha)$ is also the expected fraction of nodes with balance exactly α or $1-\alpha$, so by the definition of $\Delta(\alpha)$ we have

$$\Delta(\alpha) = f^-(\alpha) - f^+(\alpha), \quad 0 < \alpha < 1/2. \tag{9}$$

Since $b_n = 0$ for all n when α is irrational, we have $\Delta(\alpha) = 0$ for irrational α, i.e.,

$$f^-(\alpha) = f^+(\alpha),$$

and the continuity of f at irrationals follows. Furthermore, since (8) gives $\Delta(\alpha) > 0$ for rational α, (9) proves the discontinuity of f at rational α.

Having verified property (iv), it remains to verify (iii): Let $0 < \alpha < \beta < 1/2$. Then there is a rational $\gamma \in (\alpha, \beta)$ and, by (8), $\Delta(\gamma) > 0$. Together with the monotonicity of f this gives

$$f(\alpha) \geqslant f^-(\gamma) > f^+(\gamma) \geqslant f(\beta),$$

as desired.

Table 1 gives values of $f(\alpha)$ and $\Delta(\alpha)$ for $0 \leqslant \alpha \leqslant 1/2$ in increments of $1/60$; Fig. 3 shows the

TABLE 1

α	$f(\alpha)$	$\Delta(\alpha)$
$0 = 0.00000$	1.00000	0.00000
$1/60 = 0.01667$	0.99910	0.00002
$1/30 = 0.03333$	0.99643	0.00018
$1/20 = 0.05000$	0.99206	0.00060
$1/15 = 0.06667$	0.98605	0.00143
$1/12 = 0.08333$	0.97845	0.00280
$1/10 = 0.10000$	0.96931	0.00485
$7/60 = 0.11667$	0.95490	0.00002
$2/15 = 0.13333$	0.94203	0.00143
$3/20 = 0.15000$	0.92389	0.00060
$1/6 = 0.16667$	0.91831	0.02281
$11/60 = 0.18333$	0.88865	0.00002
$1/5 = 0.20000$	0.88473	0.03985
$13/60 = 0.21667$	0.84026	0.00002
$7/30 = 0.23333$	0.82884	0.00018
$1/4 = 0.25000$	0.82512	0.07944
$4/15 = 0.26667$	0.74265	0.00143
$17/60 = 0.28333$	0.73435	0.00002
$3/10 = 0.30000$	0.71697	0.00485
$19/60 = 0.31667$	0.70589	0.00002
$1/3 = 0.33333$	0.70322	0.19722
$7/20 = 0.35000$	0.50353	0.00060
$11/30 = 0.36667$	0.49461	0.00018
$23/60 = 0.38333$	0.48170	0.00002
$2/5 = 0.40000$	0.47643	0.03985
$5/12 = 0.41667$	0.43315	0.00280
$13/30 = 0.43333$	0.41343	0.00018
$9/20 = 0.45000$	0.40302	0.00060
$7/15 = 0.46667$	0.39438	0.00143
$29/60 = 0.48333$	0.38814	0.00002
$1/2 = 0.50000$	0.38630	0.38630

Values of the functions $f(\alpha)$ and $\Delta(\alpha)$, $0 \leqslant \alpha \leqslant 1/2$, in increments of $1/60$.

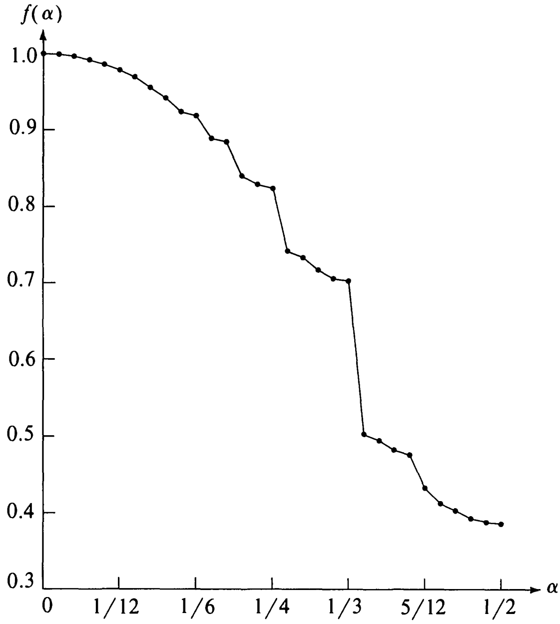


FIG. 3. The function $f(\alpha)$ plotted from the values given in the table.

values of $f(\alpha)$ in a graphical form. Since $\Delta(\alpha)$ is the expected fraction of nodes with balance exactly α or $1-\alpha$, it follows that $\Delta(1/2)=f(1/2)$ because the range $[1/2, 1/2]$ consists only of the number $1/2$. Also, $\Delta(0)=0$ since no node can have balance of 0.

Acknowledgment. The authors are grateful to Kenneth J. Supowit, Tak-Wai Chan, and the referees for their comments.

References

1. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.
2. D. E. Knuth, The Art of Computer Programming: Volume 3, Sorting and Searching, Addison-Wesley, Reading, Mass., 1972.
3. E. M. Reingold, J. Nievergelt, and N. Deo, Combinatorial Algorithms: Theory and Practice, Prentice-Hall, Englewood Cliffs, N.J., 1977.
4. M. Tenenbaum and H. Pollard, Ordinary Differential Equations, Harper and Row, New York, 1963.

A LITTLE COLOR IN ABSTRACT ALGEBRA

G. P. WENE

Division of Mathematics, Computer Science, and Systems Design, The University of Texas at San Antonio, TX 78285

Theoretical physics continues to be a source of abstract algebraic structures that illustrate basic concepts of modern algebra. In fact, the Jordan algebras, which have been so useful in mathematics, first appeared in a 1934 paper entitled "On an Algebraic Generalization of the Quantum Mechanical Formalism" (Jordan, von Neumann, and Wigner [7]). Some other well-known algebraic structures used in theoretical physics are the rotation groups and the Lie groups. Recently an "Algebra of Color" has been discovered that illustrates many interesting properties from modern algebra.

Domokos and Kovesi-Domokos [5] in their study of the color symmetry encountered in

working with the Gell-Man quark model have determined a multiplication table for an algebra describing that symmetry.

For each quark there exists an antiparticle with exactly the opposite properties (an antiquark); the quarks and antiquarks combine to form particles known as hadrons. A meson is a hadron that is made up of a quark and an antiquark. Baryons consist of three quarks bonded together; there are also antibaryons that combine three antiquarks. The inadequacies of the original (three) quark hypotheses led to the idea of color; that is, each quark comes in three varieties. Color has proved to be an exceptionally fruitful concept in explaining the interactions of quarks, but it is a purely internal property that cannot be observed. The mesons, baryons and antibaryons do not have color; yet they are observable and can be detected.

The Construction. We now construct the algebra of color as an algebra over the real or complex numbers. First it must have seven basal elements; one for each color of quark and antiquark, and a colorless element, which we shall denote by E, the hadronic observable.

Remember: color is internal, it is not an observable. Our multiplication table has basal elements E, U_1, U_2, U_3, V_1, V_2, and V_3.

I. Mesons must be observables. A meson is a quark-antiquark pair, is observable, and hence colorless. Further,

$$U_iV_j = V_iU_j = \delta_{ij}E$$

where δ_{ij} is the Kronecker delta function.

II. Baryons (antibaryons) must be observables. Since a baryon (antibaryon) consists of three quarks (antiquarks) that combine to form a colorless observable, we must have

$$U_iU_j = \epsilon_{ijk}V_k \quad \text{and} \quad V_iV_j = \epsilon_{ijk}U_k$$

where ϵ_{ijk} is the totally antisymmetric unit tensor.

III. Finally, an observable and a quark (or antiquark) must carry the same color as the quark (antiquark) does. This gives

$$EU_i = U_iE = U_i \quad \text{and} \quad EV_i = V_iE = V_i.$$

In summary, we give the multiplication table:

TABLE 1

	E	U_1	U_2	U_3	V_1	V_2	V_3
E	E	U_1	U_2	U_3	V_1	V_2	V_3
U_1	U_1	0	V_3	$-V_2$	E	0	0
U_2	U_2	$-V_3$	0	V_1	0	E	0
U_3	U_3	V_2	V_1	0	0	0	E
V_1	V_1	E	0	0	0	U_3	$-U_2$
V_2	V_2	0	E	0	$-U_3$	0	U_1
V_3	V_3	0	0	E	U_2	$-U_1$	0

Properties of the Algebra. It is easy to see that the algebra is not associative. Other basic properties of algebras are more difficult to determine. Is the algebra flexible, that is, is the relation $(xy)x - x(yx) = 0$ true for all x and y in the algebra? Define a new product $x \circ y$ by

$$x \circ y = (xy + yx)/2 \tag{1}$$

where juxtaposition denotes the original product. Is the algebra a Jordan algebra under this new product? (If it is a Jordan algebra, it is said to be Jordan-admissible. If it is a simple Jordan

algebra under this new product, it is called J-simple.)

Wene described the algebra of color in [10] as follows:

THEOREM. *The algebra of Table* 1 *is a flexible, quadratic, J-simple algebra.*

Using Table 1, direct computation shows that the algebra is quadratic, that is, each element x satisfies an equation

$$x^2 - 2t(x)x + q(x)E = 0$$

where $t(x)$ and $q(x)$ are scalars. The theory of quadratic algebras is highly developed. (See, for example, Braun and Koecher [3].) A quadratic algebra is necessarily power associative and Jordan-admissible. If the trace, $t(x)$, that appears above is associative, then the algebra is flexible. Since $t[(xy)z - x(yz)] = 0$, the algebra is flexible. Since the norm, $q(x)$, is nondegenerate, the algebra is Jordan-admissible and simple under the product (1). This class of algebras appears in the literature as early as 1948 in Albert [2]. Table 1 easily generalizes to higher dimensions.

Comments. Interested readers are referred to Schafer [9] for information on nonassociative algebras in general and McCrimmon [8] for additional information on Jordan algebras. The mathematical formulation of the color symmetry is discussed further in Adler [1], Cvitanovic, Gonsalves, and Neville [4], and Domokos and Kövesi-Domokos [6].

References

1. S. L. Adler, Classical algebraic chromodynamics, Phys. Rev. D17, (1978) N 12, 3212–3224.
2. A. A. Albert, Power-associative rings, Trans. Amer. Math. Soc., 64 (1948) 552–593.
3. H. Braun and M. Koecher, Jordan-Algebren, Springer-Verlag, Berlin, Germany, 1966.
4. P. Cvitanovic, R. J. Gonsalves, and D. E. Neville, color-charge algebras in Adler's chromodynamics, Phys. Rev. D18, (1978) N 10, 3881–3884.
5. G. Domokos and S. Kövesi-Domokos, The algebra of color, J. Math. Phys., 19 (6) (1978) 1477–1481.
6. G. Domokos and S. Kövesi-Domokos, Towards an algebraic quantum chromodynamics, Phys. Rev. D19, (1979) N 10, 2984–2996.
7. P. Jordan, J. Von Neumann, and E. Wigner, On an algebraic generalization of the quantum mechanical formalism, Ann. of Math., (2) 36 (1934) 29–64.
8. K. McCrimmon, Jordan algebras and their applications, Bull. Amer. Math. Soc., 84 (1978) 612–627.
9. R. D. Schafer, Introduction to Nonassociative Algebras, Academic Press, New York, 1966.
10. G. P. Wene, An example of a flexible, Jordan-admissible algebra of current use in hadron physics, Hadronic J., 1 (1978) 944–954.

MISCELLANEA

79. The reluctance with which the Theory of Sets was allowed to take its individual place in the scheme of mathematics was but a symptom of the same spirit which, in the domain of scientific studies generally, is dictating the demand to relegate mathematics to the rank of the handmaid of the other sciences.... This policy, if not reversed, can only lead to such a weakening of the centre of the intellectual army as will ultimately secure its defeat: the tendency, to which it properly belongs, to restrict the period of study, is one which cannot now be too actively opposed. We should rather be citing the length of studies in the Middle Ages, as an incentive to our contemporaries to show a like devotion to the far more interesting and extensive science of to-day. A lifetime was too short to run through the gamut of knowledge in those days, and at an age when our young men expect to be speaking *ex cathedra* to students not ten years their juniors, the corresponding devotee of those far-off times was eagerly working as the servant of some venerable sage, so as to gather some of his jealously guarded stores of knowledge.

—W. H. Young, *Proc. London Math. Soc.*, (2) 24 (1925) 425.

SOLUTIONS OF THE PROBLEMS IN § 5 OF "RADON INVERSION—VARIATIONS ON A THEME," BY ROBERT S. STRICHARTZ

A. The dual transform is easily seen to be

$$R^*g(x) = \sum_{x \in y} g(y)$$

and so

$$R^*Rf(x) = \sum_{x' \in X} n(x, x')f(x')$$

where $n(x, x')$ is the number of lines containing x and x'. But $n(x, x') = 1$ if $x' \neq x$, while $n(x, x) = n(x)$, the number of lines containing x. Since we are assuming there are at least two lines, $n(x) > 1$ for every x, so

$$R^*Rf(x) = (n(x) - 1)f(x) + \sum_{x' \in X} f(x')$$

or

$$f(x) = \frac{1}{n(x) - 1} R^*Rf(x) - \frac{1}{n(x) - 1} \sum f(x') \tag{6.1}$$

which will invert Rf once we express $\sum f(x')$ in terms of Rf. One way to do this is to sum equation (6.1) over all points x and solve,

$$\sum f(x') = \left(1 + \sum \frac{1}{n(x) - 1}\right)^{-1} \sum \frac{1}{n(x) - 1} R^*Rf(x).$$

Another way is to consider the sum $\sum_{y \in Y}(n(y) - 1)Rf(y)$ where $n(y)$ denotes the number of points in y. If you substitute the definition of $Rf(y)$ and interchange the order of the sums, this becomes

$$\sum_{x \in X}\left(\sum_{x \in y} (n(y) - 1)\right)f(x).$$

But for each fixed x, $\sum_{x \in y}(n(y) - 1)$ counts all the points not equal to x exactly once, so it is just $N - 1$ where N is the number of points. Thus

$$\sum f(x) = \frac{1}{N - 1} \sum (n(y) - 1)Rf(y)$$

for another solution.

For the inversion of the k-plane transform, it is convenient to consider instead of R^* the following transformation: if g is a function on X_k and u denotes any set of k distinct points, $Tg(u) = \sum_{x_k \in \tilde{u}} g(x_k)$ where $\tilde{u}$ is the set of all x_k in X_k which contain all the points of u. Since $k + 1$ points determine a unique k-plane, there is a unique k-plane in $\tilde{u}$ containing each point not in u, so

$$\begin{aligned} TR_kf(u) &= \sum_{x \notin u} f(x) + n(\tilde{u}) \sum_{x \in u} f(x) \\ &= \sum_{x \in X_0} f(x) + (n(\tilde{u}) - 1) \sum_{x \in u} f(x) \end{aligned}$$

where $n(\tilde{u})$ is the number of k-planes in $\tilde{u}$. Since there are at least two k-planes we have $n(\tilde{u}) > 1$, so

$$\sum_{x \in u} f(x) = \frac{1}{n(\tilde{u}) - 1} TR_kf(u) - \frac{1}{n(\tilde{u}) - 1} \sum_{x \in X_0} f(x). \tag{6.2}$$

By summing (6.2) over all sets u containing a fixed point x_0 we obtain

$$\binom{N-1}{k-1}f(x_0)+\binom{N-2}{k-2}\sum_{x\neq x_0}f(x)=\sum_{x_0\in u}\left[\frac{1}{n(\tilde{u})-1}TR_kf(u)-\frac{1}{n(\tilde{u})-1}\sum_{x\in X_0}f(x)\right]. \tag{6.3}$$

Since $N>k$ the combinatorial coefficients are different, and we can solve (6.3) for $f(x_0)$ in terms of R_kf and $\Sigma_{x\in X_0}f(x)$. Finally we can express $\Sigma f(x)$ in terms of R_kf by summing (6.3) over all x_0 and solving.

B. Fix a point (k_0, m_0) in L, and let $\{y_j\}$ be any sequence of lines passing through (k_0, m_0) such that all the other points of L on y_j have distance from the origin exceeding λ_j and $\lim_{j\to\infty}\lambda_j=+\infty$. (For example, take $y_j=\{(k,m):(k-k_0)=j(m-m_0)\}$; if (k,m) is on y_j and $m\neq m_0$, then $|k-k_0|\geq j$ so $|k|\geq j-|k_0|$.) Then $\lim_{j\to\infty}Rf(y_j)=f(k_0,m_0)$ gives the inversion, since

$$|Rf(y_j)-f(k_0,m_0)|\leq\sum|f(k,m)|$$

where the sum extends over all points whose distance to the origin exceeds λ_j, and this tends to zero as $\lambda_j\to\infty$ since f is absolutely summable.

To construct the counterexample $f\not\equiv 0$ but $Rf\equiv 0$, let $y_1, y_2, \ldots$ be any enumeration without repetitions of all the lines under consideration—they are countable since any two lattice points determine a unique line. It is clear from the geometry that no line is contained in a finite union of other lines. Then define f inductively as follows: choose two points on y_1 where f will be $+1$ and -1, and make f zero elsewhere on y_1. Having defined f for all points on $y_1,\ldots,y_{n-1}$, define f on all the points of y_n where it is not already defined, making it zero on all but one of these, and choosing its value on that one so as to make $Rf(y_n)=0$. Note that after any finite stage, f is nonzero at only a finite number of points, so the sum defining Rf is finite for any line.

C. Consider first how to find $f(1)$ from Rf. Since $f(1)$ occurs only in the sum for $Rf(1)$, any inversion formula must begin $f(1)=Rf(1)+\cdots$. Since $f(2)$ occurs in the sum $Rf(1)$ and again only in the sum $Rf(2)$, we must have $R(1)=Rf(1)-Rf(2)+\cdots$. Continuing in this way, considering the values $f(n)$ in turn, it is clear that if there is an inversion formula it must be

$$f(1)=c_1Rf(1)+c_2Rf(2)+\cdots+c_nRf(n)+\cdots$$

for certain coefficients c_n that are uniquely determined. A little trial and error suggests that these coefficients must be equal to the Möbius function $\mu(n)$, which is defined to be $(-1)^k$ if n has k distinct prime factors, and 0 if n is divisible by a square of a prime. An elementary exercise in number theory shows that $\Sigma\mu(d)=0$ where the sum extends over all divisors of a fixed $n>1$ (with the convention that 1 is a divisor and $\mu(1)=1$), and this is exactly what is needed to show

$$\sum_{k=1}^{n}\mu(k)Rf(k)=f(1)+\sum_{j=n+1}^{\infty}a_jf(j)$$

where $a_j=\Sigma\mu(d)$, the sum extending over all divisors of j with $d\leq n$. Now a very crude estimate for a_j is $|a_j|\leq j$ independent of n, from which we obtain the convergence

$$\lim_{n\to\infty}\sum_{n+1}^{\infty}a_jf(j)=0\quad\text{if }|f(n)|\leq cn^{-2-\varepsilon},$$

hence $f(1)=\Sigma_{k=1}^{\infty}\mu(k)Rf(k)$. Applying the same reasoning to the sequence $f(nk)$ leads to the full inversion formula $f(n)=\Sigma_{k=1}^{\infty}\mu(k)Rf(nk)$. More careful estimates for a_j will lead to the validity

of the inversion formula under less stringent conditions on the rate of decay of $f(n)$. We do not know if it is valid for all absolutely summable functions.

D. The finite lines on T^2 may be parametrized with parameter $t, 0 \leqslant t < 1$, as $(x, y) = (x_0 + tp, y_0 + tq)$ where p and q are relatively prime integers, and (x_0, y_0) is a point on T^2. Each finite line has infinitely many such parametric representations, and while it is not hard to adopt conventions so that each line has a unique representation, it would be counterproductive to do this. Indeed we will be interested in all lines passing through a fixed point (x_0, y_0), and these all have the above representation with the same values of (x_0, y_0). Let P denote the set of all pairs (p, q) of relatively prime integers with $p \geqslant 0$ and the following conventions: if $p = 0$, then $q = 1$, if $q = 0$, then $p = 1$, and ± 1 is considered relatively prime to every integer. In this way all finite lines passing through (x_0, y_0) are in one-to-one correspondence with P via the above parametric representation.

Now the Radon transform can be written

$$Rf(x_0, y_0, p, q) = \int_0^1 f(x_0 + tp, y_0 + tq)\, dt$$

in terms of this notation. It seems tempting to define a dual transform as

$$R^*g(x_0, y_0) = \sum_{(p,q) \in P} g(x_0, y_0, p, q)$$

even though it seems unlikely that the sum defining R^*Rf will converge. Nevertheless let's compute what happens if we take $f(x, y) = e^{2\pi i(jx+ky)}$, since these are the functions that enter into the Fourier series expansion of an arbitrary function. Now

$$\begin{aligned} Rf(x_0, y_0, p, q) &= \int_0^1 e^{2\pi i(j(x_0+tp)+k(y_0+tq))}\, dt \\ &= e^{2\pi i(jx_0+ky_0)} \int_0^1 e^{2\pi it(jp+kq)}\, dt \\ &= \delta(jp + kq) f(x_0, y_0) \end{aligned}$$

since the integral is zero unless $jp + kq = 0$. Now if $(j, k) \neq (0,0)$ then there is exactly one pair $(p, q) \in P$ with $jp + kq = 0$, so

$$\sum_{(p,q) \in P} Rf(x_0, y_0, p, q) = f(x_0, y_0)$$

for such f. On the other hand if $(j, k) = (0,0)$, i.e., $f \equiv 1$, then the series diverges. However the constant term in the Fourier series is the mean value of the function, which is expressible in terms of Rf by

$$\int_0^1 \int_0^1 f(x, y)\, dx\, dy = \int_0^1 Rf(s, 0, 0, 1)\, ds.$$

Thus if

$$f(x, y) = \sum\sum a_{jk} e^{2\pi i(jx+ky)}$$

for a finite sum then

$$f(x_0, y_0) = \int_0^1 Rf(s, 0, 0, 1)\, ds + \sum_{(p,q) \in P} \left[Rf(x_0, y_0, p, q) - \int_0^1 Rf(s, 0, 0, 1)\, ds \right]. \tag{6.4}$$

(Here we have used the obvious fact that $R(f - c) = Rf - c$ for any constant c.) The last sum will also be finite, since the bracketed term will be nonzero only for those $(p, q) \in P$ correspond-

ing to (j, k) appearing in the Fourier series for f.

More generally we can look at the convergence of the inversion formula (6.4) when f has an infinite Fourier series. If Q is any finite subset of P, let $\tilde{Q}$ denote the (usually infinite) subset of the lattice consisting of all (j, k) for which there exist nonzero $(p, q) \in Q$ with $jp + kq = 0$, and $(0,0) \in \tilde{Q}$ if and only if $(0,0) \in Q$. Then by the same reasoning that leads to (6.4) we have

$$\int_0^1 Rf(s,0,0,1)\,ds + \sum_{(p,q)\in Q}\left[Rf(x,y,p,q) - \int_0^1 Rf(s,0,0,1)\,ds\right] = \sum_{(j,k)\in\tilde{Q}} a_{jk}e^{2\pi i(jx+ky)}. \tag{6.5}$$

The left side of (6.5) is a finite sum and so is always well-defined, but the right side is an infinite sum. If we assume that f has an absolutely convergent Fourier series, then (6.5) is valid pointwise and we can take the limit as $Q \to P$ in any manner, thus obtaining (6.4) where the convergence is absolute and uniform. If we only assume f is continuous, then (6.5) is only valid in the mean-square, and taking the L^2 limit we obtain (6.4) in the L^2 sense.

E. By the method of Section 4 we have to compute the Fourier transform of the characteristic function of one square, say

$$S_0 = \{(x_1, x_2) : |x_1| \leq \tfrac{1}{2} \quad \text{and} \quad |x_2| \leq \tfrac{1}{2}\},$$

and show that it does not vanish on any circle. But the Fourier transform is just

$$\int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{-ix_1\xi_1}e^{-ix_2\xi_2}\,dx_1\,dx_2 = 4\xi_1^{-1}\xi_2^{-1}\sin\tfrac{1}{2}\xi_1\sin\tfrac{1}{2}\xi_2$$

which vanishes only on the "grid" $\xi_1 = 2\pi k$ or $\xi_2 = 2\pi k$ for k a nonzero integer, and of course the grid contains no circles.

If we look at the boundaries of squares, the Fourier transform we have to look at is the line integral of $e^{-ix_1\xi_1}e^{-ix_2\xi_2}$ around the square, which is easily computed to be

$$-4\xi_1^{-1}\sin\tfrac{1}{2}\xi_1\cos\tfrac{1}{2}\xi_2 - 4\xi_2^{-1}\sin\tfrac{1}{2}\xi_2\cos\tfrac{1}{2}\xi_1.$$

Now this vanishes exactly when

$$\xi_1^{-1}\tan\tfrac{1}{2}\xi_1 = -\xi_2^{-1}\tan\tfrac{1}{2}\xi_2,$$

so this can vanish identically on a circle of radius $2r$ only if the identity

$$\tan(r\sin\theta) = -\tan\theta\tan(r\cos\theta) \tag{6.6}$$

holds for all θ. Clearly this is preposterous, but to actually give a proof of its preposterousness takes a little work. Here is one proof: first set $\theta = \pi/4$, so $\tan(\sqrt{2}r/2) = 0$ or $r = 2\pi k/\sqrt{2}$ for some integer k. Substituting this back in (6.6) we would have

$$\tan\left(\frac{2\pi k}{\sqrt{2}}\sin\theta\right) = -\tan\theta\tan\left(\frac{2\pi k}{\sqrt{2}}\cos\theta\right).$$

But by choosing θ so that $\sin\theta = \sqrt{2}/4$, $\cos\theta = \sqrt{14}/4$, $\tan\theta = 1/\sqrt{7}$, we are led to

$$\tan\left(\pi\frac{k}{2}\right) = \frac{1}{\sqrt{7}}\tan\left(\frac{\pi k\sqrt{7}}{2}\right)$$

which is impossible since the left side is either 0 or ∞ while the right side is neither since $\sqrt{7}$ is irrational.

Finally in the case of the vertices the Fourier transform is $4\cos\frac{1}{2}\xi_1, \cos\frac{1}{2}\xi_2$ which again vanishes only on a grid.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by November 30, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2950. *Proposed by Ko-Wei Lih, Academia Sinica.*

The inner side of a semicircle (including diameter) is a mirror. A light ray emitting from the zenith makes an angle α with the vertical line, $0 \leqslant \alpha \leqslant \pi/2$. Characterize α such that the light ray will hit the zenith after finitely many reflections. (The context can be enlarged to a hemispherical mirror, including horizontal base.)

E 2951. *Proposed by Barbara Turner, California State University, Long Beach.*

Let $d_k(n) = \Sigma_{j=1}^{n}(2j-1)^k$, the sum of the kth powers of the first n odd positive integers, for $k \geqslant 0$, $n \geqslant 1$. Show that if p and q are to be relatively prime, $p > q > 0$, the only nontrivial solution which satisfies $[d_k(n)]^p = [d_m(n)]^q$ for all n is the identity

$$\left[\sum_{j=1}^{n}(2j-1)^0\right]^2 = \left[\sum_{j=1}^{n}(2j-1)^1\right]^1.$$

E 2952. *Proposed by A. McD. Mercer, University of Guelph.*

Prove that

$$\sin\theta \geqslant \frac{2}{\pi}\theta + \frac{\theta}{12\pi}(\pi^2 - 4\theta^2) \quad \text{in} \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}.$$

E 2953. *Proposed by John J. Wahl, Mt. Pocono, PA.*

Let A, B, X, Y be variables subject to the condition $AX - BY = 1$.

(a) Find explicit polynomials u and v in A, B, X, Y with integer coefficients such that $A^4u - B^4v = 1$.

(b) Prove in fact that for any positive m and n there are u and v such that $A^m u - B^n v = 1$.

E 2954. *Proposed by A. Meir, University of Alberta, Edmonton, Canada.*

Let $f(n)$ be a nonnegative real valued function defined for every natural number n. Suppose $f(a+b) \leqslant f(a) + f(b)$ whenever $|b-a| \leqslant \kappa$ (κ fixed). Does $\lim_n f(n)/n$ exist? (See E 2841 [1980, 577].)

E 2955. *Proposed by Greg Huber, M.I.T., Cambridge, MA.*

Let T be a torus in R^3 obtained by rotating the circle $(x-2)^2 + y^2 = 1$ in the xy plane about the y axis. Let π be a plane tangent to the torus T at the point $(0,0,1)$. Find the volume and surface area of the small region obtained by slicing T with π.

SOLUTIONS OF ELEMENTARY PROBLEMS

Elementary Symmetric Functions of $1, 2, \ldots, p-1$

E 2861 [1980, 823]. *Proposed by George Shulman, Teaneck, NJ.*

Let $p > 3$ be a prime; A_l is the lth elementary symmetric function of the set $\{1, 2, \ldots, p-1\}$. If l is odd, $1 < l < p$, prove $A_l \equiv 0 \bmod p^2$. (Wolstenholme's theorem is the case $l = p-2$.) Can the relation $A_l \equiv 0 \bmod p^2$ hold if l is even?

Solution by D. M. Bloom, Brooklyn College; R. Breusch, Amherst, MA; K. Brown, Fort Lauderdale, FL; L. L. Foster, California State University, Northridge; W. Johnson, Bowdoin College; L. Kuipers, Mollens Vs, Switzerland; O. P. Lossers, Eindhoven Institute of Technology, Netherlands; D. E. Orr, Universtiy of South Alabama; Nan-Shan Shou, University of Hong Kong.

Define $\quad f(X) = \Pi_{i=1}^{p-1}(X-i) = X^{p-1} + \Sigma_{l=1}^{p-2}(-1)^l A_l X^{p-1-l} + (p-1)!$.

Now the polynomial $X^{p-1} - 1$ has (mod p) just the zeros $1, 2, \ldots, p-1$, by Fermat's little theorem, and by the theorem of unique factorization over a field. Thus $f(X) \equiv X^{p-1} - 1 \bmod p$, and $A_l \equiv 0 \bmod p$. Since f is unchanged if $\Pi(X-i)$ is replaced by $\Pi(X-(p-i))$, the relation $f(X) = f(-X+p)$ follows; thus $f(-Y) = f(Y+p)$, so that

$$(*) \qquad Y^{p-1} + \Sigma_{l=1}^{p-2} A_l Y^{p-1-l} = (Y+p)^{p-1} + \Sigma_{l=1}^{p-2}(-1)^l A_l (Y+p)^{p-1-l}.$$

Using $A_l \equiv 0 \bmod p$ if $1 \leqslant l \leqslant p-2$, (*) shows that if l is odd, $1 < l < p$, then $A_l \equiv 0 \bmod p^2$.

Finally, reduction of (*) mod p^3 leads to

$$A_l (Y+p)^{p-1-l} = A_l Y^{p-1-l} - A_l (l+1) p Y^{p-2-l}$$

for $1 \leqslant l \leqslant p-2$. Thus from (*)

$$A_l \equiv (-1)^l A_l + (-1)^l p l A_{l-1} \bmod p^3 \quad (3 \leqslant l \leqslant p-2).$$

Hence for odd l, $1 < l < p$, the conditions $A_{l-1} \equiv 0 \bmod p^2$ and $A_l \equiv 0 \bmod p^3$ are equivalent.

If $C_k := \Sigma_1^{p-1} i^k$, then (i) $p^2 \mid C_k \Leftrightarrow p^2 \mid A_k$; (ii) if k is odd, $1 < k < p$, then $p^2 \mid C_k$; (iii) if k is even, $k < p-1$, then $p^2 \mid C_k \Leftrightarrow p^3 \mid C_{k+1}$.

Let B_{2k} be the $(2k)$th Bernoulli number. Then $p^2 \mid A_{2k} \Leftrightarrow p \mid B_{2k}$. Examples are $(k,p) =$ (8, 3617), (9, 43867), (10, 283), (10, 617), (11, 131), (11, 593), (12, 103), (12, 2294797), (16, 37), (6, 691).

Foster referred to Dickson's *History*, vol. 1, p. 100. Kuipers called attention to his article *Der Wolstenholmesche Satz*, Elemente der Math., 35 (1980) 62-64, where his proofs use Taylor's theorem.

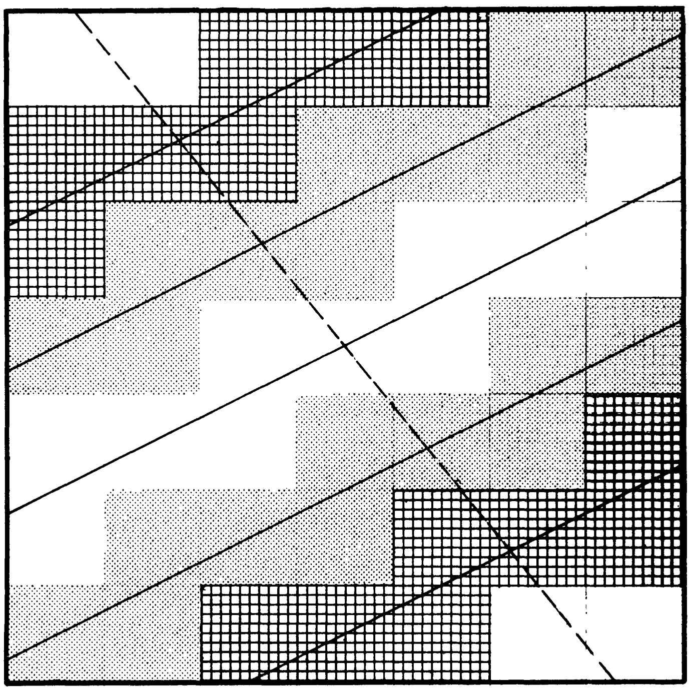
Covering an $n \times n$ Chessboard with $n-1$ Lines

E 2862 [1980, 823]. *Proposed by T. Keller, Honolulu, Hawaii.*

For $n \geqslant 3$, show that $n-1$ straight lines are sufficient to go through the interior of every

square of an $n \times n$ chessboard. *Are $n - 1$ lines necessary?

Solution to part 1 by M. Goldberg, Washington, D. C., Henry Osner, Modesto Junior College, and University of Hartford Problem Group. The corners of the board are $(-n, n)$, $(0, n)$, $(-n, 0)$, $(0,0)$. Draw $n-2$ lines L_i with slope $1/2$, the first intersecting the base line at $(-9/2, 0)$, and the rest spaced so that the vertical distance between lines is $6/4$ units. L_i has equation $y = \frac{1}{2}x + 9/4 + 6i/4$ $(i = 0, 1, \ldots, n-3)$. (L_{i+1} is obtained by shifting L_i one unit to the left and then one unit up.) L_1 goes through the points $(-9/2, 0)$, $(-4, 1/4)$, $(-3, 3/4)$, $(-5/2, 1)$, $(-2, 5/4)$, $(-1, 7/4)$, $(-1/2, 2)$, $(0, 9/4)$. L_{n-2} goes through the point $(-n + 9/2, n)$. (Thus the arrangement is symmetric under 180 degree rotation.) Leaving out the two rightmost squares in the lowest row, and the two leftmost squares in the top row, every unit square has an interior point on one or another of the lines. The $(n-1)$st line can be drawn to pass through the interiors of the four orphaned squares.



Also solved by J. Dou (Spain), L. Harkleroad & E. Harkleroad, V. Hernandez (Spain), J. Hook (student), M. Josephy (Costa Rica), H. M. Marston, P. -Y. Wu, an anonymous solver, and the proposer.

Remark by the editors. For $m \leqslant n$, let $f(m, n)$ be the minimum number of lines needed to cover an $m \times n$ board, in the sense that some interior point of every unit square is on one or another of the lines. Clearly $f(1, n) = 1$, $f(2, n) = 2$, $f(n-1, n) \leqslant n - 1$. It can be shown that $f(3,4) = 3$, $f(4,5) = 4$, $f(5,6) = 5$. Thus $f(3, n) = 3$, $f(4, n) = 4$, $f(5, n) = 5$. If (*) $f(n-1, n) = n-1$, then $f(m, n)$ is determined for all m, n: $f(n-k, n) = n-k$, if $0 < k < n$, $n \geqslant 3$; $f(n, n) = n - 1$. Some inductive argument is needed to complete the proof of (*).

The integrals $\int_0^\infty [\ln(x/(x+a))]^p dx$

E 2865 [1981, 66]. *Proposed by David K. Cohoon, USAF School of Aerospace Medicine, San Antonio, Texas.*

Show that for positive real a

$$\int_0^\infty \left[\ln\left(\frac{x}{x+a}\right)\right]^2 dx = \pi^2 a/3.$$

Solution by Fergus J. Gaines, University College, Dublin, Ireland, and W. W. Zachary, Naval Research Laboratory. Note $a > 0$. More generally, let $s > 1$ be real. Set $I(a, s) = \int_0^\infty (\log y)^s dx$, $y = x/(x + a)$. Make the substitution $y = e^{-t}$. Thus

$$I(a, s) = (-1)^{s+1} a \int_0^\infty t^s d(e^t - 1)^{-1} = (-1)^s as \int_0^\infty dt\, t^{s-1}/(e^t - 1),$$

using integration by parts. This gives $I(a, s) = (-1)^s as\Gamma(s)\zeta(s)$. (See Whittaker and Watson, *Modern Analysis*, p. 266.) If $s = 2N$, $I(a, 2N) = a2^{2N-1}\pi^{2N} | B_{2N} |$ (*ibid.* pp. 268, 269).

Klinger referred to CRC tables. Ehlers, Foster, and Kappus referred to formula 4.261.5 from Gradstein and Ryzik's *Tables*. Some solvers used contour integration, some computed $\partial I(a, s)/\partial a$, some integrated a power series. Some used l'Hôpital's rule. Rangan noted the implicit assumption $a > 0$.

Also solved by M. Ashbaugh, K. L. Bernstein, I. Bivens, D. Bode (West Germany), A. Boghossian & J. Wichman (Saudi Arabia), R. Bournas, J. N. Boyd, D. W. Brown, P. S. Bruckman, P. F. Byrd, J. E. Chance, P. Chauveheid (Belgium), Chico Problem Group, G. Costa, E. Deutsch, P. F. Ehlers, E. A. Enneking, A. Facchini (Italy), Martie Fields (student), L. L. Foster, L. Gesing (Hong Kong), S. H. Greene, J. L. Hafner, J. R. Hatcher, R. Heller, V. Hernandez (Spain), J. Holland, E. Johnston, O. Jørsboe (Denmark), H. Kappus (Germany), I. N. Katz, P. Khageh-Khalili, M. S. Klamkin, K. A. Klinger, E. L. Koh (Saudi Arabia), V. Konecny, A. R. Kräuter (Austria), L. Kuipers (Switzerland), I. E. Leonard (Canada), R. A. Leslie, O. P. Lossers (The Netherlands), A. Lyzzaik (Saudi Arabia), B. Margolis (France), N. Martin (Canada), D. Moore, A. M. Nadel, R. B. Nelson, C. P. Rangan (India), B. Ross, F. Safier, St. Olaf Problem Solving Group, I. A. Sakmar (Canada), C. W. Schelin, H. D. Shane, B. L. R. Shawyer (Canada), D. Shelupsky, S. Simanca (student, Venezuela), F. C. Smith, A. Stenger, R. A. Struble, A. E. Trojanowski, E. Trost (Switzerland), D. Tyler, H. R. van der Vaart, L. Van Hamme (Belgium), S. K. Venkatesan (India), M. Vowe (Switzerland), W. V. Webb, J. Wiener, R. L. Young, and the proposer.

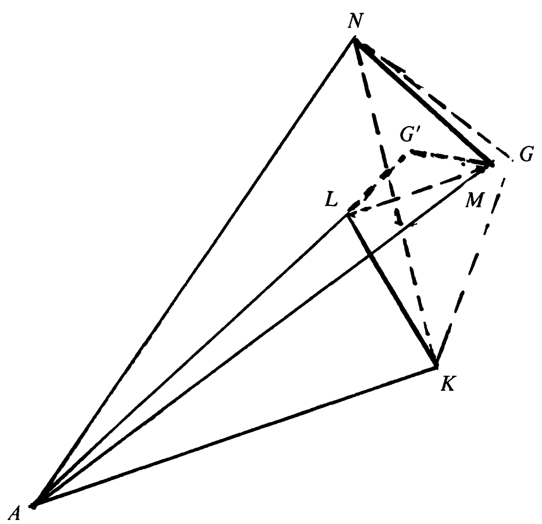
Similarly Oriented Equilateral Triangles

E 2866 [1981, 66]. *Proposed by Jordi Dou, University of Barcelona, Spain.*

Let AKL, AMN be equilateral triangles. Prove that the equilateral triangles LMX, NKY are concentric (if Y is on the properly chosen side of NK).

Solution by Clayton W. Dodge, University of Maine at Orono. The theorem is a special case of the following theorem:

Let AKL and AMN be directly similar isosceles triangles with apex angles of measure $\theta°$ and both at A. Let GNK and $G'LM$ be directly similar isosceles triangles with apex angles at G and at G', each of measure $180° - \theta°$, and each oriented in the same sense as AKL. Then G and G' coincide.



The theorem is seen to be self-converse, and the problem is the special case $\theta° = 60°$. To prove the theorem, note that the rotation through angle $\theta°$ about A carries triangle AMK to triangle ANL, so KM and LN are equal in length and meet at angles of $\theta°$ and $180° - \theta°$. Let B be the point of intersection of the perpendicular bisectors of LM and NK. (If these bisectors coincide, point B is readily located on that line to satisfy the conditions that follow.) Then $BL = BM$ and $BN = BK$, so triangles BLN and BMK are directly congruent. Hence a rotation about B through angle $180° - \theta°$ carries BLN to BMK. Hence BLM and BNK are directly similar isosceles triangles of apex angle $180° - \theta°$. Hence $G = G' = B$. □

Comparative Variation of Two Real Functions

E 2867 [1981, 66]. *Proposed by Dennis K. Mick, Carroll College, Waukesha, Wisconsin.*

Let $h(t) > 0$, $g(t) > 0$ be continuous functions for $0 \leqslant t < \infty$, $\int_0^\infty h(t)\,dt = \infty$, $\int_0^\infty g(t)\,dt < \infty$. Prove that there exist arbitrarily large values of r such that for any s satisfying $0 \leqslant s \leqslant r$ we have $\int_{r-s}^{r+s} h(t)\,dt \geqslant \int_{r-s}^{r+s} g(t)\,dt$.

I. *Solution by David M. Wells, Pennsylvania State University.* It is in fact possible to prove a stronger statement:

PROPOSITION. *Under the states hypotheses there exist arbitrarily large values of r such that for any s satisfying* $0 < s \leqslant r$ *we have*

$$\int_{r-s}^{r} h(t)\,dt > \int_{r-s}^{r} g(t)\,dt \quad \text{and} \quad \int_{r}^{r+s} h(t)\,dt > \int_{r}^{r+s} g(t)\,dt.$$

Proof. Let $H(x) = \int_0^x h(t)\,dt$, $G(x) = \int_0^x g(t)\,dt$ and $F(x) = G(x) - H(x)$. Then $F(x)$, $G(x)$, $H(x)$ are continuous, $G(x)$ and $H(x)$ are increasing, $F(0) = G(0) = H(0) = 0$, $\lim_{x\to\infty} F(x) = \lim_{x\to\infty} H(x) = \infty$ and $\lim_{x\to\infty} G(x) = L < \infty$. We may choose an increasing sequence $\{x_0, x_1, \ldots\}$ with $\lim_{n\to\infty} x_n = \infty$ and $F(x_n) = 2nL$. Since the total variation of $G(x)$ is L, there is some $r_n \in (x_{n-1}, x_n)$ for which $F^{-1}[F(r_n)] = \{r_n\}$ for $n \geqslant 1$. In particular there are arbitrarily large values of r for which $F^{-1}[F(r)] = \{r\}$. Furthermore if $r > 0$ has this property, the stated conditions of $F(x)$ imply that $F(r - s) < F(r) < F(r + s)$ for $0 < s \leqslant r$. This in turn implies the conclusion of the proposition.

II. *Solution by Bruce L. Montgomery, student, Carnegie-Mellon University.* We shall prove the theorem assuming that g and h satisfy the weaker condition of integrability on $[0, t]$ for any $t \geqslant 0$. The proof is by contradiction. If the conclusion is false, then for each r greater than some $M > 0$, there exists an $s(r)$, $0 \leqslant s(r) \leqslant r$, such that

$$\int_{r-s(r)}^{r+s(r)} g(t)\,dt > \int_{r-s(r)}^{r+s(r)} h(t)\,dt.$$

Note, in fact, for this strict inequality to hold, we must have $0 < s(r) \leqslant r$.

If $h(t) > 0$, $g(t) > 0$ are integrable functions on $[0, t]$ for any $t \geqslant 0$, and $\int_0^\infty h(t)\,dt = \infty$, $\int_0^\infty g(t)\,dt < \infty$, then there exists $N > M$ such that (*)$\int_M^N h(t)\,dt > 2\int_0^\infty g(t)\,dt$. Clearly, $\{(r - s(r), r + s(r)\colon r \in [M, N]\}$ is an open covering for $[M, N]$. Thus, finitely many $I_j = [r_j - s(r_j), r_j + s(r_j)]$ cover $[M, N]$, and we may assume that if $E = \cup_j I_j$, then each $t \in E$ is contained in at most two of the I_j. Then

$$2\int_0^\infty g(t)\,dt \geqslant 2\int_E g(t)\,dt \geqslant \sum_j \int_{I_j} g(t)\,dt > \sum_j \int_{I_j} h(t)\,dt \geqslant \int_E h(t)\,dt \geqslant \int_M^N h(t)\,dt$$

which contradicts (*).

Also solved by O. P. Lossers (Netherlands), N. Martin (Canada), M. D. Meyerson, D. Moore, S. Noltie, D. A. Rawsthorne, C. R. Rosentrater, A. Smuckler (Israel), J. Suck (Germany), J. Tripp, and the proposer.

Suck noted that "arbitrarily large values of r" cannot be replaced by "all r."

$G/Z(G) \simeq H/Z(H)$; $G = HZ(G)$

E 2869 [1981, 147]. *Proposed by Desmond MacHale, Cork, Ireland.*

Let G be a group with center $Z(G)$ such that $[G\colon Z(G)]$ is finite. Prove that if H is a subgroup of G then $[G\colon Z(G)] \geqslant [H\colon Z(H)]$. Find necessary and sufficient conditions for equality and show that, in the case of equality, $G/Z(G) \simeq H/Z(H)$ and $G' = H'$ where prime denotes the commutator subgroup.

Solution by Anders Bager, Hjørring, Denmark. Since $H \cap Z(G) \leqslant Z(H)$, we have the following sequence:

$$HZ(G)/Z(G) \overset{\alpha}{\to} H/(H \cap Z(G)) \overset{\beta}{\to} H/Z(H) \overset{\gamma}{\to} I(H),$$

where $I(H)$ is the set of inner automorphisms of H. Also α, γ are isomorphisms, and β is an epimorphism (homomorphism onto). But also, $G \geqslant HZ(G)$, so that $I(G) \simeq G/Z(G) \geqslant HZ(G)/Z(G)$; thus $|I(G)| \geqslant |I(H)|$.

Suppose now that $[G\colon Z(G)] = [H\colon Z(H)]$, finite. Then α is an isomorphism, hence $G/Z(G) \cong I(H) \cong H/Z(H)$. Let $g \in G$. Then $\gamma_g = \eta_h$ with $h \in H$, hence $gh^{-1} = z \in Z(G)$. Thus $G = HZ(G)$. Again without a finiteness condition, it is easily seen that $G = HZ(G)$ is sufficient for $[G\colon Z(G)] = [H\colon Z(H)]$. Since central elements cancel out of commutators, under this condition we have $G' \leqslant H'$, hence $G' = H'$.

Remark by Paul L. Chabot, California State University at Los Angeles. If $|G\colon Z(G)|$ is not finite, then $Z(G) = Z(H) = 1$, $G \simeq H$, $G \neq H$ can hold simultaneously (thus $G \neq HZ(G)$). Counterexample: G is the "countable restricted alternating group" = "directed limit of $\mathfrak{A}_n$" = set of all even permutations of the natural numbers with finite support. H is the subgroup of G that fixes "1."

Richard Montgomery, student, University of California at Berkeley stated the condition in the form: every coset of $Z(G)$ in G intersects H.

Richards noted that if $[G\colon ZG] = [G'\colon ZG']$, $G \neq 1$, then G is not solvable. Flanigan sent an extensive analysis.

Also solved by D. W. Brown, T. Bu (Norway), D. C. Buchtal, P. L. Chabot, J. A. Cuenca (Spain), H. M. Edgar, L. Erlebach, L. L. Foster, F. J. Flanigan, J. Hook (student), Dinh Thế Hùng, R. D. Hurwitz, L. Jones, S. C. King, R. D. Konyndyk, Leung Gesing (Hong Kong), B. Glastad, C. Libis, C. Lyons, M. R. Modak (India), R. Montgomery, W. Myers, V. Pambuccian (student, Rumania), R. G. E. Pinch (England), I. M. Richards (England), E. F. Schmeichel, E. J. Taft, J. Teitelbaum, University of Hartford Problem Group, E. Veed, G. R. Walls, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by November 30, 1982. The solver's full post-office address should be on each sheet.

6392. *Proposed by C. Ward Henson, Bruce Reznick and Lee A. Rubel, University of Illinois.*

Is there some $x_0 > 0$ such that if $f(x) = x^x$, then $f^{(n)}(x) \geqslant 0$ for all $x \geqslant x_0$ and all $n = 0, 1, 2, \ldots$?

SOLUTIONS OF ADVANCED PROBLEMS

An Improper Integral from a Singular Distribution

6320 [1980, 759]. *Proposed by Edgar A. Cohen, Jr., Naval Surface Weapons Center, Silver Spring, Maryland.*

The principal value integral

$$P=\lim_{T\to\infty}\frac{1}{\pi^2}\int_{-T}^{T}\int_{-T}^{T}\frac{\sin ht}{t}\,\frac{\sin ku}{u}\,\frac{\sin(t+u)}{t+u}\,dt\,du,\quad 0\leqslant h\leqslant 1,\quad 0\leqslant k\leqslant 1,\tag{1}$$

arises in probability theory, where it furnishes a representation for the cumulative distribution function whose mass is concentrated on the line $y=x$ [**1**, p. 123 $(a=0,\ b=0]$:

$$F(x,y)=\begin{cases}x, & x\leqslant y\\ y, & y\leqslant x\end{cases}\tag{2}$$

where $(x,y)\in[0,1]\times[0,1]$. Therefore, one sees that $P=\min(h,k)$. Is it true that the iterated integral

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\sin ht}{t}\,\frac{\sin ku}{u}\,\frac{\sin(t+u)}{t+u}\,dt\,du\tag{3}$$

likewise exists? Can one evaluate (1) directly? If (3) exists, can it be obtained from *ab initio* considerations?

1. Fisz, Marek, Probability Theory and Mathematical Statistics, 3rd ed., Wiley, New York, 1963.

Solution by M. L. Glasser, Clarkson College, Potsdam, NY.

$$P=\lim_{T\to\infty}\frac{1}{\pi^2}\int_0^1 dy\int_{-T}^{T}dt\int_{-T}^{T}du\frac{\sin ht}{t}\,\frac{\sin ku}{u}\cos(u+t)y.$$

By symmetry, this is

$$=\lim_{T\to\infty}\frac{1}{\pi^2}\int_0^1 dy\int_{-T}^{T}dt\frac{\sin ht}{t}\cos ty\int_{-T}^{T}du\frac{\sin ku}{u}\cos uy.$$

Now it is known that [Erdelyi et al., *Tables of Integral Transforms*, vol. 1, McGraw-Hill, 1954, p. 18, 1.6 (1)]

$$\lim_{T\to\infty}\int_{-T}^{T}dt\frac{\sin at}{t}\cos ty=\pi\theta(a-y)$$

where $\theta(x)=\begin{cases}1 & x>0\\ 0 & x<0\end{cases}$; thus

$$P=\int_0^1 dy\,\theta(h-y)\theta(k-y)=\min(h,k).$$

Also solved by H. Guggenheimer, L. Kuipers (Switzerland), and the proposer.

A Condition on Minimal Left Ideals in a Ring

6324 [1980, 826]. *Proposed by I. N. Herstein, University of Chicago.*

Let R be a ring (not necessarily with 1) having no nilpotent ideals. Suppose that $L\neq 0$ is a left ideal in R such that Ra is a minimal left ideal of R for every $a\neq 0$ in L. Prove that L itself must be a minimal left ideal of R.

Solution by Joel K. Haack, Oklahoma State University. Let $0\neq a\in L$; then Ra is a minimal left ideal. Since R has no nilpotent ideals, [**1**, Lemma 1.3.1] applies to give an idempotent $e\in Ra\subseteq L$ with $Ra=Re$. Since $0\neq e=e^2\in Le$, Le is a nonzero left ideal of R. Thus by the minimality of Re, $Re=Le$. Decompose L as the direct sum $L=Le+\{l-le\mid l\in L\}$. We will show that $\{l-le\mid l\in L\}=0$. For if not, let $b\in L$ with $0\neq b-be$. Then $R(b-be)$ is a minimal left ideal, so as above, there exists an idempotent $f\in L$ with

$$Rf=R(b-be)\subseteq\{l-le\mid l\in L\}.$$

Since

$$f=0+f=fe+(f-fe)\in Le+Rf,$$

we see that $fe = 0$. Let $g = e + f - ef$; then $eg = e$ and $fg = f$. Since $g \in L$ and $g \neq 0$, Rg is a minimal left ideal. But $0 \neq Rf \subseteq Rg$ and $Rf \neq Rg$, for $rf = rfg \in Rg$, but $e = eg \in Rg \setminus Rf$. Thus $\{l - le \mid l \in L\} = 0$, so $L = Le = Ra$ is a minimal left ideal.

Reference

1. I. N. Herstein, Noncommutative Rings, Carus Mathematical Monograph No. 15, Mathematical Association of America, Washington DC, 1968.

Also solved by R. K. Amayo, Efraim P. Armendáriz, William D. Blair, F. J. Flanigan, Enzo R. Gentile (Argentina), Kwangil Koh, W. G. Leavitt, Barbara L. Osofsky, Harry F. Smith, J. Zelmanowitz, and the proposer.

A Congruence on Euler Numbers

6325* [1980, 826]. *Proposed by Barry J. Powell, Kirkland, Washington.*

Let $E_0 = 1$, $E_2 = -1$, $E_4 = 5$, $E_6 = -61$, $E_8 = 1385$, $E_{10} = -50521, \ldots,$ be the set of Euler numbers defined by $\sec x = \sum_0^\infty E_n x^n/n!$ ($|x| < \pi/2$). Prove or disprove that, for any prime $p \equiv 1$ mod 4, $E_{(p-1)/2} \not\equiv 0 \bmod p$. (It is true for $p \equiv 5 \bmod 8$. See E. Lehmer, *On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson*, Annals of Math., 39 (1938) 350-360.)

Solution by Reijo Ernvall, University of Turku, Finland. We show that $E_{(p-1)/2} \not\equiv 0 \bmod p$ for all $p \equiv 1 \bmod 4$.

Proof. In [**2**, p. 37] and [**3**, p. 354] it is proved that $E_{(p-1)/2} \equiv 4\sum_1^u a^{(p-1)/2} \bmod p$, where $u = (p-1)/4$. It is well known that $a^{(p-1)/2} \equiv (a/p) \bmod p$, where (a/p) denotes the Legendre symbol. Hence $E_{(p-1)/2} \equiv 0 \bmod p$ iff $\sum_1^u (a/p) = 0$. But as a consequence of Dirichlet's class number theorem we know that $\sum_1^u (a/p) > 0$ (see e.g. [**1**]). The assertion is proved.

An elementary proof for our statement in the case $p \equiv 5 \bmod 8$ is also given in [**2**, p. 339].

References

1. B. C. Berndt, Classical theorems on quadratic residues, Enseign. Math., 22 (1976) 261–304.

2. R. Ernvall, Generalized Bernoulli numbers, generalized irregular primes, and class number, Ann. Univ. Turku. Ser. A I, 178 (1979) 72.

3. E. Lehmer, On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson, Ann. of Math., 39 (1938) 350–360.

THE AMERICAN

MATHEMATICAL MONTHLY

Volume 89, Number 7 August-September 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

ON SOME ABSTRACT PROPERTIES OF BINOMIAL COEFFICIENTS

DEAN S. CLARK
Department of Mathematics, Hofstra University, Hempstead, NY 11550

0. Introduction. In (1.7) below we present an "abstract" combinatorial identity, conceptually no more difficult than the great classic which heads the following list, and which includes as special cases:

(a) Newton's binomial formula $\Sigma_j \binom{n}{j} x^j y^{n-j} = (x+y)^n$

(b) Vandermonde's identities

$$\sum_j \binom{n}{j}\binom{m}{i-j} = \binom{m+n}{i}; \quad \sum_j \binom{n}{j}\binom{m}{i+j} = \binom{m+n}{m-i}; \quad \sum_j \binom{n}{j}\binom{m-j}{i}(-1)^j$$

$$= \binom{m-n}{m-i}; \quad \sum_j \binom{n}{j}\binom{m+j}{i}(-1)^j = (-1)^n \binom{m}{i-n}$$

(c) The "Fibonacci convolutions"

$$\sum_j \binom{n}{j} F_{m-j} = F_{m+n}; \quad \sum_j \binom{n}{j} 2^{n-j} F_{m-j} = F_{m+2n}; \quad \sum_j \binom{n}{j} F_{m+j} = F_{m+2n};$$

$$\sum_j \binom{n}{j} 2^j F_{m+j} = F_{m+3n},$$

where $\{F_n\}_{n>-\infty}$ is any real sequence satisfying $F_{n+2} = F_{n+1} + F_n$ (e.g., the Fibonacci numbers)

(d) The Gaussian formula

$$\sum_{j=0}^{n} \prod_{i=1}^{k-1} (j+i) = \frac{1}{k} \prod_{i=1}^{k} (n+i)$$

and its generalization

$$\sum_{j=0}^{n} \prod_{i=1}^{k} (n-j+i) \prod_{l=1}^{q} (j+l) = \binom{n+k+q+1}{n} k!q!$$

(e) The lacunary sums

$$\sum_j \binom{n}{2j} = 2^{n-1}; \quad \sum_j \binom{3m+1}{3j} = \frac{2^{3m+1} + (-1)^m}{3};$$

$$\sum_j \binom{8m+2}{4j} = 2^{8m}; \quad \sum_j \binom{10m+1}{5j} = \frac{2^{10m+1} + f_{10m+2} + f_{10m}}{5}$$

where $\{f_n\}_{n \geqslant 0} = \{1, 1, 2, \dots\}$ is the Fibonacci sequence;

$$\sum_j \binom{12m+3}{6j} = \frac{2^{12m+2} - 1}{3}; \dots$$

(f) $\Sigma_j \binom{n}{j}(-1)^j = 0$ and its generalizations:

The author taught political economy at Roger Williams College, Bristol, Rhode Island, 1970–74, and was also on the Fine Arts faculty, teaching courses in jazz improvisation. He was research associate in the Division of Applied Mathematics, Brown University, 1977–78; lecturer at the Universität Zürich, Institut für Operations Research, 1978–81; and joined the Hofstra faculty this fall. His research has been mainly in ordinary differential equations and applied probability. He wrote this paper as an avid amateur combinatorialist.

$$\sum_j \binom{n}{j}(-1)^{[j/2]} = 0 \text{ or } \pm 2^{[(n+1)/2]};$$

$$\sum_j \binom{n}{j}(-1)^{[j/3]} = 0, \qquad \pm 2 \cdot 3^{[(n-1)/2]}, \quad \text{or } \pm 4 \cdot 3^{[(n-1)/2]}; \ldots$$

where $[x]$ is the greatest integer not greater than x

(g) $\Sigma_j \binom{n}{j} j = n2^{n-1}$ and its generalizations:

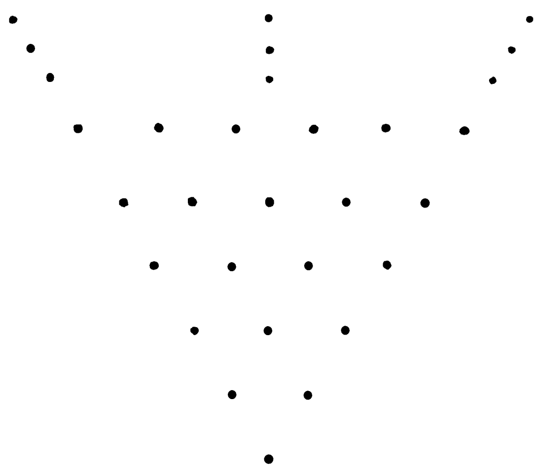
$$\sum_j \binom{n}{j}[j/2] = (n-1)2^{n-2}, n \geqslant 1; \qquad \sum_j \binom{3m+2}{j}[j/3] = m2^{3m+1};$$

$$\sum_j \binom{n}{j}[j/2](-1)^j = 2^{n-2}, n \geqslant 2; \qquad \sum_j \binom{6m}{j}[j/3](-1)^j = (-1)^m 3^{3m-1}, \qquad m \geqslant 1; \ldots$$

$$\vdots$$

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
\cdot & & & & \cdot & & & & \cdot \\
\cdot & & & & \cdot & & & & \cdot \\
\cdot & & & & \cdot & & & & \cdot
\end{array}$$

(a)



(b)

$$\begin{array}{l}
(1+x)^0 \times 1 \ \ ^{(+)}\ x \quad x^2 \quad x^3 \quad x^4 \quad \ldots \quad x^n \\
\quad (1+x)^1 \times \searrow 1 \swarrow \quad x \quad x^2 \quad x^3 \quad \cdots \quad x^{n-1} \\
\quad\quad (1+x)^2 \times 1 \quad x \quad x^2 \ \ldots \ x^{n-2} \\
\quad\quad\quad \ddots \qquad\qquad \iddots \\
\quad\quad\quad\quad (1+x)^{n-1} \times 1 \quad x \\
\quad\quad\quad\quad\quad (1+x)^n \times 1
\end{array}$$

(c)

FIG. (1.1)

The list goes on. Furthermore, the point of view which enables one to see how all these identities are unified results, in our opinion, is an enhancement of computational insight and power beyond the classical methods. This is illustrated by an application which follows in Section 2. To establish

notation, define the binomial coefficients $\binom{n}{j}$ to be those real numbers which satisfy

$$\binom{0}{j} = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{0.1a}$$

and

$$\binom{n+1}{j} = \binom{n}{j} + \binom{n}{j-1}, \tag{0.1b}$$

where n and j are, respectively, nonnegative and arbitrary integers. Thus, the expression $\Sigma_j\binom{n}{j}c_j$ makes sense when the summation is over all integer j-values.

1. The Binomial Theorem, Once Again. Consider the following "proof" of Newton's classical identity $\Sigma_{j=0}^{n}\binom{n}{j}x^j = (1+x)^n$: Take Pascal's triangle, shown below in Fig. (1.1a), extract the physical configuration only, and turn it upside down, (b). Let the terms x^j be the new initial values, and operate on them as though *they* were binomial coefficients, (c). In other words, to go from the zeroth row to the first row, write the sum $x^j + x^{j+1} = (1+x)x^j$ below and between the two terms x^j and x^{j+1}, exactly as one does in Pascal's triangle. The common multiplier $(1+x)$ is factored out at the left margin. Continue the process until a single term—the correct answer—appears at the bottom of the triangle. This represents, quite plainly, a genuine proof, but why does it work? One answer is that we want to add up the x^j terms in such a way that x^j is counted $\binom{n}{j}$ times in the sum. The reader can verify that this is precisely what happens as the summation process works its way downward in the array of Fig. (1.1c). But here is a more intriguing answer: The objects $(1+x)^n x^j \equiv \left\{ {n \atop j} \right\}$ *are*, in a sense, "binomial coefficients." They satisfy the relation

$$\left\{ {n+1 \atop j} \right\} = \left\{ {n \atop j} \right\} + \left\{ {n \atop j+1} \right\}, \tag{1.1}$$

and it is really this property which is critical in Fig. (1.1c). In fact, if $\left\{ {n \atop j} \right\}$, $n \geqslant 0$, $j > -\infty$, are *any* numbers which satisfy (1.1), then the same counting argument or simple induction yields the identity

$$\sum_j \binom{n}{j}\left\{ {0 \atop j} \right\} = \left\{ {n \atop 0} \right\}. \tag{1.2}$$

There is a certain symbolic inevitability about (1.2), and its impact would be all the greater if such objects $\left\{ {n \atop j} \right\}$ existed in abundance. They do. But, before the demonstration, we need to reach for even greater generality than (1.2), and this requires discussion of another motivating example. This is the problem of finding a closed-form expression for the sums $\Sigma_j\binom{n}{3j+q}$, $q = 0, 1, 2$ (which, incidentally, has appeared a number of times in the MONTHLY; see [1] for the references). We do it by inclusion-exclusion of the row sums of Pascal's triangle. For the sake of simplicity, let $n = 3m$, $q = 0$. In Fig. (1.2a), below, the elements in the triangle to be added are circled in the nth row, $n = 6$. Their antecedents as defined by (0.1) are indicated in the $n-1$st row. The first step is to add (include) the *entire* $n-1$st row sum, 2^{n-1}, watchful of the two elements which make this sum too large, $\binom{5}{1} = \binom{5}{4} = 5$, (b). Subtract (exclude) these, (c), by subtracting the entire $n-2$nd row sum, 2^{n-2}. But this excludes too much, so $\binom{4}{2} = 6$ is put back by including the entire $n-3$rd row sum, and so forth. The complete procedure is shown in the schematic of Fig.

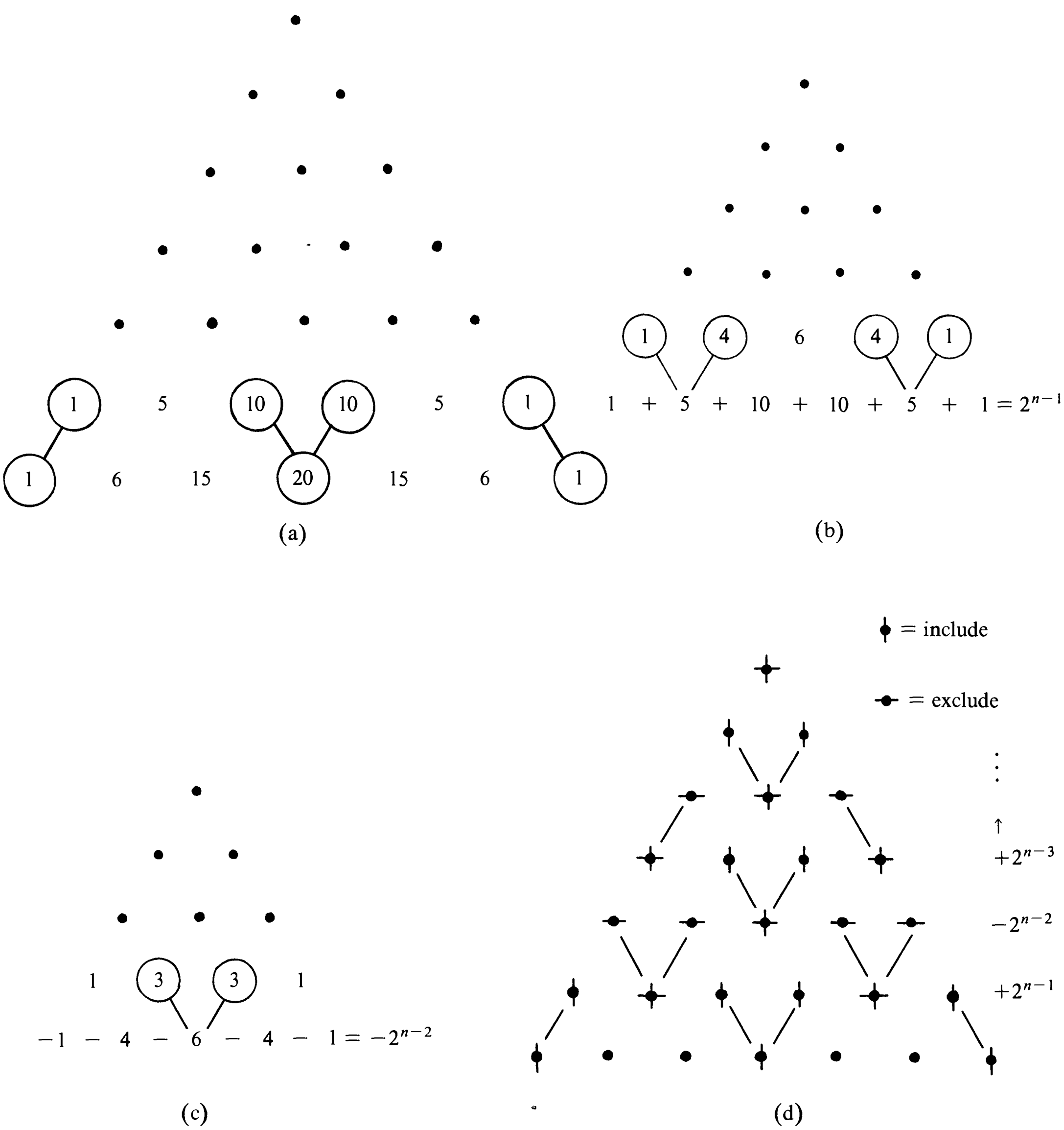


FIG. (1.2)

(1.2d), where only the positions of the binomial coefficients are shown. A position with a vertical (horizontal) line through it means that the corresponding binomial coefficient has been included (excluded) in the sum. The process leads to the much simpler problem of computing a telescoping geometric sum, i.e., if n is a multiple of 3

$$\sum_j \binom{n}{3j} = (-1)^n + \sum_{j=1}^{n} (-1)^{j-1} 2^{n-j}. \tag{1.3}$$

In general,

$$\sum_j \binom{n}{3j+q} = \begin{cases} (-1)^n & \text{if } n = 3m - q \\ 0 & \text{otherwise} \end{cases} + \frac{2^n + (-1)^{n+1}}{3}, \tag{1.4}$$

where the second term on the r.h.s. of (1.4) is the collapsed remnant of the geometric sum in (1.3). Now generalize what happened in (1.3). Let the "primal" problem be: Find a closed formula for

$$\text{(P)} \qquad \sum_j \binom{n}{j} c_j, \ \{c_j\}_{j>-\infty} \text{ an arbitrary real sequence,} \tag{1.5a}$$

the "dual" problem: Compute

$$\text{(D)} \qquad (-1)^n \alpha_n + \sum_{j=1}^{n} (-1)^{j-1} 2^{n-j} \lambda_j, \qquad \alpha_n, \lambda_j \text{ real numbers.} \tag{1.5b}$$

The λ_j tell us how many times to include or exclude the $(n-j)$th row sum of Pascal's triangle, with α_n the residual to be added or subtracted at the end. The fundamental question—when are the primal and dual equal?—has the following answer.

THEOREM (1.1). (P) = (D), *with* $c_j = \left\langle {0 \atop i+j} \right\rangle$ *and* $\alpha_n = \left\langle {n \atop i} \right\rangle$ *for any fixed integer* i, *if and only if*

$$\left\langle {n+1 \atop j} \right\rangle = \lambda_{n+1} - \left\langle {n \atop j} \right\rangle - \left\langle {n \atop j+1} \right\rangle \quad \textit{for all } n \geqslant 0,\ j > -\infty. \tag{1.6}$$

A further generalization:

$$\sum_j \binom{n}{j} \left\langle {r \atop i+j} \right\rangle = (-1)^n \left\langle {r+n \atop i} \right\rangle + \sum_{j=1}^{n} (-1)^{j-1} 2^{n-j} \lambda_{r+j} \tag{1.7}$$

for any $n, r \geqslant 0, i > -\infty$, *if and only if* (1.6) *holds.*

Proof. For the sufficiency, (1.6) $\Rightarrow$ (1.7), write the equivalent version of (1.7)

$$\left\langle {r+n \atop i} \right\rangle = (-1)^n \sum_j \binom{n}{j} \left\langle {r \atop i+j} \right\rangle + \sum_{j=1}^{n} (-1)^{n+j} 2^{n-j} \lambda_{r+j}, \tag{1.8}$$

and use

$$\left\langle {r+n+1 \atop i} \right\rangle = \lambda_{r+n+1} - \left\langle {r+n \atop i} \right\rangle - \left\langle {r+n \atop i+1} \right\rangle$$

with (1.8) to complete an induction argument on n. For the necessity, (1.7) $\Rightarrow$ (1.6), simply set $n = 1$ in (1.8). □

(1.7) is the first-mentioned identity, which specializes to those listed in the Introduction for various choices of initial condition $R^0 \equiv \left(\left\langle {0 \atop j} \right\rangle \right)_{j>-\infty}$ and sequence $\Lambda \equiv \{\lambda_n\}_{n \geqslant 1}$. Clearly, something quite interesting is going on here: First we pretended that the multipliers x^j were the initial elements in a kind of upside-down Pascal triangle, and got the correct answer—certainly no coincidence—at the end. Then we considered another "inverse process" in Pascal's triangle, a counting process which suggested a problem dual to that of computing $\Sigma_j \binom{n}{j} c_j$. Equality of the primal and dual forces our attention to the objects denoted $\left\langle {n \atop j} \right\rangle$, which must obey a recursion, (1.6), reminiscent of that of the classical binomial coefficients, but with a difference. The initial value R^0 is free for us to choose, and the auxiliary variable λ_n adds an extra degree of freedom. Here is the precise relationship between the new coefficients $\left\langle {n \atop j} \right\rangle$ and the classical ones: If we define

$$\begin{bmatrix} n \\ j \end{bmatrix} \equiv (-1)^n \left\langle {n \atop -j} \right\rangle \quad \text{and} \quad \varepsilon_n \equiv (-1)^{n+1} \lambda_{n+1}, \tag{1.9}$$

it follows that

$$\begin{bmatrix} n+1 \\ j \end{bmatrix} = \begin{bmatrix} n \\ j \end{bmatrix} + \begin{bmatrix} n \\ j-1 \end{bmatrix} + \varepsilon_n, \qquad n \geqslant 0, j > -\infty. \tag{1.10}$$

When $\varepsilon_n \equiv 0$ and $\begin{bmatrix} 0 \\ j \end{bmatrix} = \begin{cases} 1 & j=0 \\ 0 & \text{otherwise} \end{cases}$, the $\begin{bmatrix} n \\ j \end{bmatrix}$ become the familiar binomial coefficients $\binom{n}{j}$. For a final specialization, let $\Lambda \equiv 0$ and $\left\{ n \atop j \right\} \equiv (-1)^n \left\langle n \atop j \right\rangle$, so that

$$\left\{ n+1 \atop j \right\} = \left\{ n \atop j \right\} + \left\{ n \atop j+1 \right\}, \qquad n \geqslant 0, j > -\infty. \tag{1.11}$$

With i, $r = 0$ (1.7) takes the pretty form

$$\sum_j \binom{n}{j} \left\{ 0 \atop j \right\} = \left\{ n \atop 0 \right\}. \tag{1.12}$$

We have finally come full circle: Letting

$$\left\{ n \atop j \right\} = \left(1 + \frac{x}{y}\right)^n \left(\frac{x}{y}\right)^j,$$

we see that these objects satisfy (1.11). Thus, (1.12) must hold. But

$$\left\{ 0 \atop j \right\} = \left(\frac{x}{y}\right)^j, \left\{ n \atop 0 \right\} = \left(1 + \frac{x}{y}\right)^n,$$

consequently

$$\sum_j \binom{n}{j} \left(\frac{x}{y}\right)^j = \left(1 + \frac{x}{y}\right)^n.$$

Multiplying both sides by y^n yields the binomial theorem.

Evidently, the $\left\langle n \atop j \right\rangle$ and $\left\{ n \atop j \right\}$ are entities similar to the ordinary binomial coefficients, but with a life of their own and clear statements of kinship to the latter, (1.6)-(1.12). They are more than simply formalisms, having arisen quite explicitly from a counting process in Pascal's triangle (recall the role of $\left\langle n \atop j \right\rangle$ as an inclusion-exclusion residual). The mirror imagery of (1.9), recursion (1.11), and dichotomy (1.5) make the "duality" nomenclature persuasive. We'll call them dual binomial coefficients (dbc's), and an array of $\left\langle n \atop j \right\rangle$ or $(-1)^n \left\langle n \atop j \right\rangle$, whichever is notationally more convenient, a dual (Pascal) array. (When $\begin{bmatrix} n \\ j \end{bmatrix} \equiv \binom{n}{j}$, (1.9) implies $\left\langle n \atop -j \right\rangle = (-1)^n \binom{n}{j}$. Thus, in a rectangular array with the usual initial condition (0.1a), the dbc's appear as a left-handed reflection of Pascal's (right) triangle.)

EXAMPLE (1.1). Dual binomial coefficients exist in abundance.

Letting $\left\{ n \atop j \right\}$ equal, respectively,

$$\binom{m+n}{i-j}, \quad \binom{m+n}{m-i-j}, \quad (-1)^j \binom{m-n-j}{i-n}, \quad (-1)^{n+j} \binom{m+j}{i-n},$$

$$F_{m+n-j}, 2^{-n-j} F_{m+2n-j}, F_{m+2n+j}, \quad \text{and} \quad 2^j F_{m+3n+j}$$

yields the identities of (b) and (c) of the Introduction, via (1.12). We already know that $(1+x)^n x^j$ are dbc's. Therefore, repeated differentiation or integration of $(1+x)^n x^j$ with respect to x generates more solutions of (1.11).

EXAMPLE (1.2). The elegant formula

$$\sum_{j=0}^{n} \prod_{i=1}^{k-1} (j+i) = \frac{1}{k} \prod_{i=1}^{k} (n+i)$$

was called "Gaussian" in the Introduction because of the legendary special case $k = 2$. It is an immediate result of the identity ($m = n + k$)

$$\sum_j \frac{\binom{n}{j}}{(m-j)\binom{m}{j}} = \frac{1}{m-n},$$

which is (1.12) when $\left\{ {n \atop j} \right\} \equiv \left((m-n-j)\binom{m-n}{j} \right)^{-1}$. Similarly, the impressive-looking generalization in (0.d) follows from

$$\sum_j \frac{\binom{n}{j}}{(m-r+j)\binom{m}{r-j}} = \frac{1}{(m-r)\binom{m-n}{m-r}},$$

when

$$\left\{ {n \atop j} \right\} \equiv \left((m-r+j)\binom{m-n}{m-r+j} \right)^{-1}.$$

EXAMPLE (1.3). By constructing a dual array, one can prove several identities for the price of one, and show, moreover, that certain identity pairs are naturally associated. For example,

$$\sum_j \binom{n+1}{j}(-1)^j j = \sum_j \binom{n}{j}(-1)^j = 0$$

follows from the array of Fig. (1.3), below, and (1.7). For the first sum, the first row is used as initial value, for the second, the second row.

	$j=0$	$j=1$	$\cdots$				
R^0:	0	-1	2	-3	4	$-5 \cdots$	$n=0$
	1	-1	1	-1	$1 \cdots$		$n=1$
	0	0	0	0	$0 \cdots$		$\vdots$
	0	0	$0 \cdots$				
	0	$0 \cdots$					
	$\vdots$						

FIG. (1.3): $\left\langle {n \atop j} \right\rangle$, $n, j \geq 0$, $\Lambda \equiv 0$.

$$\sum_j \binom{n+1}{j}[j/2] = \sum_j \binom{n}{j} j = n2^{n-1}$$

follows from

$$\begin{array}{l} \quad\;\; + \\ \quad\;\; \leftarrow \\ \| \downarrow \begin{array}{ccccccccc} 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 & \cdots \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots & \\ 1 & 3 & 5 & 7 & 9 & 11 & \cdots & & \\ 4 & 8 & 12 & 16 & 20 & \cdots & & & \\ \vdots & & & & & & & & \end{array} \end{array}$$

FIG. (1.4): $\left\{ {n \atop j} \right\}$, $n, j \geqslant 0$.

There are even hierarchies of identities:

$$\sum_j \binom{n+2}{j}(-1)^j[j/2] = \sum_j \binom{n+1}{2j+1} = \sum_j \binom{n}{j} = 2^n$$

follows from (1.7) ($r, i = 0$) and the single array

$$\begin{array}{ccccccccc} 0 & 0 & 1 & -1 & 2 & -2 & 3 & -3 & \cdots \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & \cdots & \\ 1 & 1 & 1 & 1 & 1 & \cdots & & & \\ 0 & 0 & 0 & 0 & \cdots & & & & \\ 0 & 0 & \cdots & & & & & & \\ 0 & \cdots & & & & & & & \\ & & & & \vdots & & & & \end{array}$$

FIG. (1.5): $(-1)^n\left\langle {n \atop j} \right\rangle$, $j, n \geqslant 0$, $\Lambda = \{0,0,2,0,0,0,\ldots\}$.

2. The Dual Pascal Array. In this section we state a "soft" theorem on the structure of one type of dual array—actually it is little more than a simple observation, and a proof is hardly required. What we will do is identify a class of permissible initial rows R^0 and a minimal requirement on the sequence Λ. For the former, a typical row of the classical Pascal triangle is used. Abstracted are its qualitative properties of nonnegativity, symmetry, and semimonotonicity. This last nonce word refers, of course, to the rising and symmetric falling of the values $\binom{n}{j}$ as j goes from zero to n. If this row is denoted by $V = (v_0, v_1, \ldots, v_{m-1})$, $m \geqslant 1$, then

$$0 \leqslant v_j \leqslant v_{j+1},\ j = 0,\ldots,[m/2]-1, \text{ and} \tag{2.1a}$$

$$\begin{cases} v_{[m/2]-j} = v_{[m/2]+j} \\ \text{or} \\ v_{[(m-1)/2]-j} = v_{[m/2]+j} \end{cases} \qquad j = 0,\ldots,[(m-1)/2] \tag{2.1b}$$

encapsulates the idea. Note that if m is odd, (2.1b) reduces to a single condition. Let the left shift operator, acting on such vectors V, be defined in the natural way: $L_\alpha V = (v_\alpha, v_{\alpha+1}, \ldots, v_{m-1}, v_0, \ldots, v_{\alpha-1})$, $0 \leqslant \alpha \leqslant m-1$. An initial value $R^0 = \left(\left\langle {0 \atop j} \right\rangle\right)_{j>-\infty}$ is called permissible if there exist integers m and α, with $m \geqslant 1$, $0 \leqslant \alpha \leqslant m-1$, such that

$$\begin{aligned} &L_\alpha\left(\left\langle {0 \atop 0} \right\rangle, \ldots, \left\langle {0 \atop m-1} \right\rangle\right) && \text{satisfies (2.1), and} \\ &\left\langle {0 \atop j} \right\rangle = \left\langle {0 \atop j+m} \right\rangle && \text{for all } j. \end{aligned} \tag{2.2}$$

A permissible initial row is therefore periodic, consisting of nonnegative *blocks*

$$b_0 \equiv \left(\left\langle {0 \atop j} \right\rangle\right)_{0 \leqslant j \leqslant m-1}$$

which, modulo a left shift, display the qualitative features of the mth row of Pascal's triangle. Let F^m denote the set of such infinite vectors. Our specification for an initial value automatically imposes a restriction on $\Lambda \equiv \{\lambda_n\}_{n\geq 1}$. If all the $\left\langle {n \atop j} \right\rangle$ are to be nonnegative ((2.1a)), then Λ must at least satisfy

$$\lambda_{n+1} \geq \max_j \left(\left\langle {n \atop j} \right\rangle + \left\langle {n \atop j+1} \right\rangle \right), \quad n \geq 0, \tag{2.3}$$

simply from the definition (1.6). Let (2.3) be the criterion of admissibility for Λ. The coarser features of the resulting dual array are given in

THEOREM (2.1). *Under conditions* (2.2) *and* (2.3), *every row R^n is permissible*, i.e., $R^n \in F^m$, $n \geq 0$. *Let b_n denote the block* $\left(\left\langle {n \atop j} \right\rangle \right)_{0 \leq j \leq m-1}$, *and say that two blocks b and b' are equivalent* (*in the sense of relative magnitude*), $b \simeq b'$, *if and only if $L_\alpha b$ and $L_\alpha b'$ both satisfy* (2.1) *for some α. Then, $b_{n+2} \simeq L_1 b_n$. In other words, if T is the mapping which takes a row of the dual array to the one below it*,

$$T\colon F^m \times \mathbb{R} \to F^m, \quad \text{and} \quad T^2|_{\text{block}} \simeq L_1. \tag{2.4}$$

EXAMPLE (2.1). Let $m = 5$, with initial block $b_0 = (1,0,0,0,0)$. Certainly this is an admissible block since $(1,0,0,0,0) = L_2(0,0,1,0,0)$. Tighten criterion (2.3) by changing the inequality to equality, i.e.,

$$\lambda_{n+1} \equiv \max_j \left(\left\langle {n \atop j} \right\rangle + \left\langle {n \atop j+1} \right\rangle \right), \quad n \geq 0.$$

By writing out the first few rows of the resulting array, the reader will see the left-shifting action of Theorem (2.1) and, incidentally, the following result: The Fibonacci numbers are dbc's. In fact, each block of the array has the form $(0, f_{n-1}, f_n, f_{n-1}, 0)$ modulo a shift, $f_n = n+1$st Fibonacci number. He or she will also be close to deriving the identities for the sums $\Sigma_j \left({n \atop 5j+q} \right)$, $q = 0,\ldots,4$, one of which is shown in (0.e).

Returning to the mapping T of (2.4), most generally $T\colon \mathbb{R}^\infty \times \mathbb{R} \to \mathbb{R}^\infty$. That is, $(T(X,\lambda))_j = \lambda - x_j - x_{j+1}$, $\lambda \in \mathbb{R}$, and $X = (\ldots, x_{-1}, x_0, x_1, \ldots) \in \mathbb{R}^\infty$. Under these very general conditions, the iterates of T produce a chaotic array. Theorem (2.1) reduces the problem, in a sense, to "finite dimensions", the natural first step in tackling an infinite dimensional problem. None of the finer properties of the resulting array are mentioned (although some inkling of them is given in Example (2.1)). Even so, the very elementary insight of Theorem (2.1) enables us to make nontrivial computations, and elegantly at that.

Application—a paradox in combinatorial probability. A game is to be played on m positions or "states" arranged in clockwise order as in Fig. (2.1) below ($m = 7$). Each state carries a money payoff $\pi_j, \Sigma_{j=0}^{m-1}\pi_j = 0$. A ball, beginning at position 0, moves successively clockwise around the states until it is stopped by an as yet undetermined random mechanism. The player receives the payoff—possibly a penalty—of the ball's final state. Now, if the stopping mechanism assures that each state has equal probability of coming up, the game is "fair" in the probability-theoretic sense. Under suitable conditions and repeated playing, we expect that the average payoff will approach zero. Consider this mechanism: A perfectly fair coin shall be tossed (independently) N times, with the ball advanced from state i to $i+1$ when heads are observed, not advanced when tails are observed. N should be a large integer. The game ends after the Nth toss. With respect to the playing field of Fig. (2.1), here are a couple of candidates for N:

$$N_1 = 7{,}984{,}650{,}129$$
$$N_2 = 6{,}045{,}235{,}176. \tag{2.5}$$

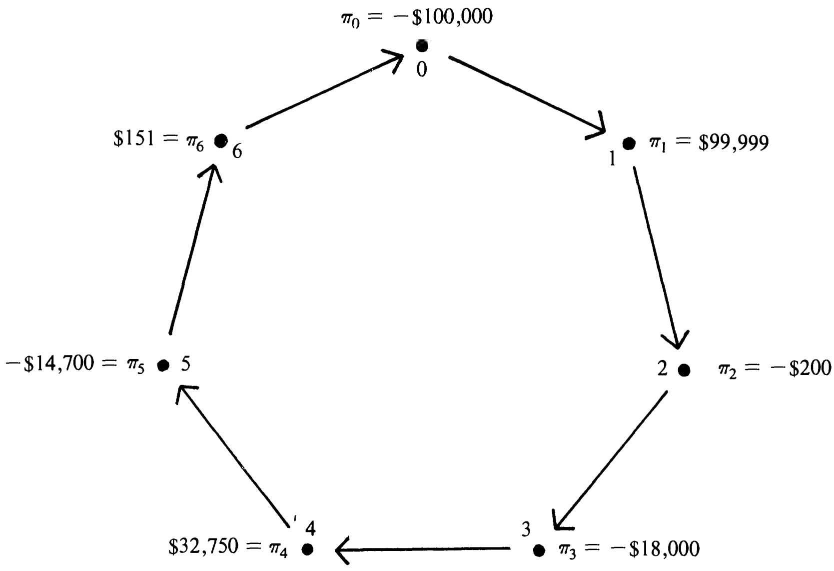


FIG. (2.1)

Is it possible that, for such N, the game is *unfair*? The fact is, if the game of Fig. (2.1) is played repeatedly with the same N, one of the two values (2.5) will, with probability one, lead to long-run profits, the other to long-run losses. But which is which? The expected value of the game—the deciding factor—is $2^{-N}\Sigma_j\binom{N}{j}\pi(j)$, where $\pi(\cdot)$ is the periodic function whose first m values are the π_j. By itself, this representation tells us nothing. The problem is purely computational:

$$\text{Is } \sum_j \binom{7{,}984{,}650{,}129}{j}\pi(j) \text{ positive or negative?} \tag{2.6}$$

How to proceed? Certainly not with a computer, since most of the nearly 8 billion terms summed in (2.6) are large enough to melt down any mere machine. One approach, appealing to the periodicity of $\pi(\cdot)$, is to call upon classical formulae for sums of the type $\Sigma_j\binom{n}{mj+q}$ (usually attributed to [2], a simple proof is in [1]). But these strike us as unsatisfying, at least in an aesthetic sense. Arising almost immediately from the relation $1 + \omega + \omega^2 + \cdots + \omega^{m-1} = 0$, ω a primitive mth root of unity, they are strictly formal, that is, they are not developed from combinatorial arguments and divert attention to fairly involved *trigonometric* expansions. In contrast, Theorem (2.1) enables a surprisingly easy answer to (2.6).

Let a dual array have initial block $b_0 = (1, 0, \ldots, 0)$, (2.3) hold, and consider sums of the form $\Sigma_j\binom{n}{mj+q}$, $q = 0, \ldots, m-1$. Because $\Sigma_{j=0}^{m-1}\pi_j = 0$, it follows from (1.7), with $r = 0$ and $i = -q$, that

$$\sum_{j=0}^{N}\binom{N}{j}\pi(j) = (-1)^N\sum_{q=0}^{m-1}\left\langle \begin{matrix} N \\ -q \end{matrix}\right\rangle \pi_q.$$

Now, fix attention on the data of (2.5) and Fig. (2.1). $N_1 = 2K_1 + 1$, where $K_1 \bmod 7 = 0$. Theorem (2.1) implies that the block b_{N_1} is equivalent, in the sense of $\simeq$, to $b_1 = (0, 1, 1, 1, 1, 1, 0)$. More precisely, b_{N_1} has the form $(c_0, c_1, c_2, c_3, c_2, c_1, c_0)$, $0 \le c_i < c_{i+1}$, $i = 0, 1, 2$. It is easily verified that the values c_i are distinct and positive (compute the first few rows of the array). The problem becomes the much simpler one of determining the sign of

$$(-1)^{N_1}(c_0(\pi_0+\pi_1)+c_1(\pi_2+\pi_6)+c_2(\pi_3+\pi_5)+c_3\pi_4)$$
$$=(-1)^{N_1}((c_1-c_0)(-\pi_0-\pi_1)+(c_2-c_1)(\pi_3+\pi_4+\pi_5)+(c_3-c_2)\pi_4).$$

Fig. (2.1) shows that $\pi_0+\pi_1<0$, $\pi_3+\pi_4+\pi_5>0$, and $\pi_4>0$. Because N_1 is odd, the answer to (2.6) is: *Negative*. N_2 is the correct choice, because $N_2=2K_2$ with $K_2 \bmod 7=4$. Thus, $b_{N_2}\simeq L_4b_0$ which again has the form $(d_0,d_1,d_2,d_3,d_2,d_1,d_0), 0\leqslant d_i<d_{i+1}, i=0,1,2$. This time, however, N_2 is even, giving the game a positive expected value.

Despite irrefutable calculations, intuition somehow demands that there be *indifference* between N_1 and N_2. After all, how can certain states be more favorable than others when the ball's ultimate destination is a random outcome awaiting the result of billions of impartial coin tosses? They can be—and *are*—and, at the same time, our intuition is in normal working order! The resolution of the paradox, as the reader may have already deduced, lies in the magnitude of the expected profit or loss: It is an amount so small as to be far beyond human sensory abilities. Even the almost preposterous upper bound $\$10^{-240,000,000}$ is a crude, vast overestimate of the expected profit or loss. Where this bound came from is another story, which may be told later.

Acknowledgments. I'd like to thank Professors Philip J. Davis and Lewis Pakula for helpful discussions; and the city of Zurich, Switzerland, for Strandbad Mythenquai, where much of the thinking for this paper took place. It could not have been done without the generosity of Professor Dr. P. Kall, Director, Institut für Operations Research der Universität Zürich.

References

1. V. E. Hoggatt, Jr. and G. L. Alexanderson, Sums of partition sets in generalized Pascal triangles I, Fibonacci Quart., April, 1976.

2. Chr. Ramus, Solution générale d'un problème d'analyse combinatoire, J. Reine Angew. Math., 11 (1834) 353–355.

EXTENDING FUNCTIONS TO INFINITESIMALS OF FINITE ORDER

DAWN FISHER

Department of Mathematics and Computer Science, James Madison University, Harrisonburg, VA 22807

When the real numbers are extended to a system which includes infinitesimals, the extended system is usually a field (see [1]–[7]). In contrast, the extensions of the real numbers discussed in this article are only commutative rings with identity, in which the infinitesimals form a decreasing sequence of ideals.

There are practical advantages to the use of finite order infinitesimals which compensate for the loss of field properties. For example, when $n=1$, we get a neat algebraic way to compute derivatives of elementary functions (see examples at the ends of sections 2 and 3). These finite order infinitesimals are very simple to construct and manipulate, and the elementary functions extend easily to them. They might be suitable for an algorithmic first course in the calculus. They could be introduced simply as numbers smaller than any real number, so small that their square is negligible (i.e., zero).

The integral as antiderivative would have its usual intuitive association with area, volume, etc. A second course, principally for mathematics majors, could introduce limits, continuity, and

---

Dawn Fisher studied functional analysis at Tulane University with B. J. Pettis. She has taught at the University of Miami, Tulane, LSU, Douglass College, the University of Pennsylvania, the University of Otago (Dunedin, New Zealand), the University of Waikato (Hamilton, New Zealand), and James Madison University. She is interested in the ways physicists use mathematics.

differentiability, and in that course one could justify the use of infinitesimals in the earlier course, and extend the concepts of derivative and integral to nonelementary functions.

1. First order algebra. If we were to adjoin to the real numbers an "infinitesimal" number ι (iota) smaller than any positive real, so small in fact that $\iota^2 = 0$, then we would expect to get a system of extended numbers $a_0 + a_1\iota$ in which

$$(a_0 + a_1\iota)(b_0 + b_1\iota) = a_0b_0 + (a_0b_1 + a_1b_0)\iota + a_1b_1\iota^2$$
$$= a_0b_0 + (a_0b_1 + a_1b_0)\iota.$$

We can build such a system be defining on $R_1 = R \times R$ (R = real numbers):

(=) $(a_0, a_1) = (b_0, b_1)$ if and only if $a_0 = b_0$ and $a_1 = b_1$;

(<) $(a_0, a_1) < (b_0, b_1)$ if and only if $a_0 < b_0$ or $a_0 = b_0$ and $a_1 < b_1$;

(+) $(a_0, a_1) + (b_0, b_1) = (a_0 + b_0, a_1 + b_1)$;

(×) $(a_0, a_1) \times (b_0, b_1) = (a_0b_0, a_0b_1 + a_1b_0)$;

(·) $r(a_0, a_1) = (r, 0)(a_0, a_1) = (ra_0, ra_1)$.

R_1 is linearly ordered, a commutative ring with identity $1 = (1,0)$ in which, for any infinitesimal $\varepsilon = a\iota$, $\varepsilon^2 = 0$, and in which the infinitesimals form a maximal ideal.

If $\iota = (0, 1)$, then $(a_0, a_1) = a_0 1 + a_1\iota$ which (just as with complex numbers) we could write $a_0 + a_1\iota$. It will, however, be more convenient to represent a typical extended real number by $x + \varepsilon$, where x is real and ε is infinitesimal.

In this notation the basic properties of R_1 are:

(=) $x + \varepsilon = x' + \varepsilon'$ if and only if $x = x'$ and $\varepsilon = \varepsilon'$;

(<) $x + \varepsilon < x' + \varepsilon'$ if and only if $x < x'$ or $x = x'$ and $\varepsilon < \varepsilon'$;

(+) $(x + \varepsilon) + (x' + \varepsilon') = (x + x') + (\varepsilon + \varepsilon')$;

(×) $(x + \varepsilon)(x' + \varepsilon') = (xx') + (x\varepsilon' + x'\varepsilon)$;

(·) $r(x + \varepsilon) = (rx) + (r\varepsilon)$

The following results are not hard to obtain:

(1) If $x \neq 0$, then $1/(x + \varepsilon) = (1/x) - (1/x^2)\varepsilon$. The reciprocal $1/\varepsilon$ does not exist, but $\varepsilon/\varepsilon' = a\iota/a'\iota = a/a'$ exists and is real if $\varepsilon' \neq 0$;

(2) If $x \neq 0$, n an integer, then $(x + \varepsilon)^n = x^n + nx^{n-1}\varepsilon$;

(3) If $x > 0$, r rational, then $(x + \varepsilon)^r = x^r + rx^{r-1}\varepsilon$.

If $|x + \varepsilon|$ is defined to be $x + \varepsilon$ if $x + \varepsilon \geqslant 0$ and $-x - \varepsilon$ if $x + \varepsilon < 0$, then:

(4) $|\varepsilon| = |a\iota| = |a|\iota$;

(5) If $x \neq 0, |x + \varepsilon| = |x| + (|x|/x)\varepsilon$;

(6) $|(x + \varepsilon)(x' + \varepsilon')| = |x + \varepsilon||x' + \varepsilon'|$;

(7) $|(x + \varepsilon) + (x' + \varepsilon')| \leqslant |x + \varepsilon| + |x' + \varepsilon'|$.

2. Extending algebraic functions to R_1. The natural extension of $f(x) = x^r$ is $f_*(x + \varepsilon) = (x + \varepsilon)^r$, and if r is rational, we have

$$(x + \varepsilon)^r = x^r + rx^{r-1}\varepsilon,$$

so

$$f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon.$$

Since algebraic functions are generated from the functions $f(x) = x^r$ (r rational) by the operations of multiplication by a real number, addition, multiplication, and composition, any

transformation $Tf = f_*$ which maps $f(x) = x^r$ onto $f_*(x + \varepsilon) = (x + \varepsilon)^r$ and which preserves each of these operations will map every algebraic function onto its natural extension. If there is such a transformation and if two different algebraic expressions determine the same function on R, they will also determine the same function on R_1.

We will show that there is such a transformation $Tf = f_*$, and that it does indeed satisfy $f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon$.

DEFINITION 1. Let $R_* = R + I$ be a commutative ring with identity, with maximal ideal I. A transformation $Tf = f_*$ of real functions into R_* functions is an *extension to* R_* if:

(i) if f is in the domain of T, and $\operatorname{dom}(f)$ is the domain of f, then the function with domain $\operatorname{dom}(f)$ and constant value 1 is in the domain of T, and its transform is the function with domain $\operatorname{dom}(f) + I$ and constant value 1;

(ii) if f is in the domain of T, the identity on $\operatorname{dom}(f)$ is in the domain of T, and its transform is the identity on $\operatorname{dom}(f) + I$;

(iii) if f is in the domain of T and r is any real number, then rf is in the domain of T and $(rf)_* = rf_*$;

(iv) if f and g and $f + g$ are in the domain of T, then $(f + g)_* = f_* + g_*$;

(v) if f and g and $f \times g$ are in the domain of T, then $(f \times g)_* = f_* \times g_*$;

(vi) if f and g and $f(g)$ are in the domain of T, then $f(g)_* = f_*(g_*)$.

DEFINITION 2. A transformation $Df = f'$ of real functions into real functions is a *differentiation* if:

(i) if f is in the domain of D, the function with domain $\operatorname{dom}(f)$ and constant value 1 is in the domain of D, and its transform is the function with domain $\operatorname{dom}(f)$ and constant value 0;

(ii) if f is in the domain of D, then the identity function on $\operatorname{dom}(f)$ is in the domain of D, and its transform is the function with domain $\operatorname{dom}(f)$ and constant value 1;

(iii) if f is in the domain of D and r is any real number, then rf is in the domain of D and $(rf)' = rf'$;

(iv) if f and g and $f + g$ are in the domain of D, then $(f + g)' = f' + g'$;

(v) if f and g and $f \times g$ are in the domain of D, then $(f \times g)' = f' \times g + f \times g'$;

(vi) if f and g and $f(g)$ are in the domain of D, then $f(g)' = f'(g) \times g'$.

THEOREM 1. *If $Df = f'$, transforms real functions into real functions, and $Tf = f_*$ transforms real functions into R_1 functions, and if $f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon$, then D is a differentiation if and only if T is an extension to R_1.*

Proof. Check extension and differentiation properties (i)–(vi).

THEOREM 2. *Let A be the set of real algebraic functions. There is one and only one differentiation $Df = f'$ with domain A, one and only one extension $Tf = f_*$ to R_1 with domain A, and $f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon$.*

Proof. Any two differentiations will agree that if $f(x) = x^r$ (r rational), then $f'(x) = rx^{r-1}$, and therefore will agree on every algebraic function. If $Df = f'$ is differentiation on A, then $f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon$ defines an extension $Tf = f_*$ to R_1. Any two extensions will agree on $f(x) = x^r$ (r rational) and therefore on every algebraic function.

We can now differentiate any algebraic function by purely algebraic methods. For example, if $f(x) = (1 + x^3)^{1/2}$, then

$$\begin{aligned} f_*(x + \varepsilon) &= \left(1 + (x + \varepsilon)^3\right)^{1/2} = \left(1 + (x^3 + 3x^3\varepsilon)\right)^{1/2} = \left((1 + x^3) + (3x^2\varepsilon)\right)^{1/2} \\ &= (1 + x^3)^{1/2} + (1/2)(1 + x^3)^{-1/2}(3x^2\varepsilon) = f(x) + f'(x)\varepsilon, \end{aligned}$$

so

$$f'(x) = (1/2)(1 + x^3)^{-1/2}(3x^2).$$

If $f(x) = 1/x$, then

$$f_*(x+\varepsilon) = 1/(x+\varepsilon) = (1/(x+\varepsilon))(x-\varepsilon)/(x-\varepsilon)$$
$$= (x-\varepsilon)/x^2 = (1/x) - (1/x^2)\varepsilon, \text{ so } f'(x) = -(1/x^2).$$

3. Extending differentiable functions to R_1. If A_1 is the set of differentiable functions, then $Df = f'$ is a differentiation on A_1, so $Tf = f_*(x+\varepsilon) = f(x) + f'(x)\varepsilon$ is an extension to R_1.

Moreover, T satisfies the *first order Correspondence Principle*, which is that if ϕ is in A_1, and $|\phi(h)| \leq M|h|$ for all sufficiently small h, then $|\phi_*(\varepsilon)| \leq M|\varepsilon|$ for all ε.

THEOREM 3. *If $Tf = f_*$ is an extension to R_1 with domain A_1 which obeys the Correspondence Principle, and f is in A_1, then*

$$f_*(x+\varepsilon) = f(x) + f'(x)\varepsilon.$$

Proof. If f is in A_1, and x is in dom(f), then $\phi(h) = f(x+h) - f(x) - f'(x)h$ is also in A_1, and $|\phi(h)| \leq M|h|$ for all $M > 0$, and all sufficiently small h, therefore $|\phi_*(\varepsilon)| \leq M|\varepsilon|$ for all $M > 0$, so

$$\phi_*(\varepsilon) = f_*(x+\varepsilon) - f(x) - f'(x)\varepsilon = 0,$$

and

$$f_*(x+\varepsilon) = f(x) + f'(x)\varepsilon.$$

For example, $|1 - \cos h| \leq h^2/2$ for all sufficiently small h, so $\cos_* \varepsilon = 1$ for all ε, and one can show similarly that $\sin_* \varepsilon = \varepsilon$, $\ln_*(1+\varepsilon) = \varepsilon$, and $e_*^{\varepsilon} = 1 + \varepsilon$. Since

$$|\cos(x+h) - (\cos x \cos h - \sin x \sin h)| = 0$$

for all h, therefore

$$\cos_*(x+\varepsilon) = \cos x \cos_* \varepsilon - \sin x \sin_* \varepsilon = \cos x - \varepsilon \sin x,$$

so $(\cos x)' = -\sin x$; similarly,

$$\sin_*(x+\varepsilon) = \sin x + \varepsilon \cos x$$

and

$$\ln_*(x+\varepsilon) = \ln_*(x(1+\varepsilon/x)) = \ln x + \ln_*\left(1 + \frac{\varepsilon}{x}\right) = \ln x + \frac{\varepsilon}{x},$$

and

$$e_*{}^{(x+\varepsilon)} = e^x e_*^{\varepsilon} = e^x(1+\varepsilon) = e^x + e^x\varepsilon.$$

Another sample calculation would be

$$\sec_*(x+\varepsilon) = 1/\cos_*(x+\varepsilon) = 1/(\cos x - \varepsilon \sin x)$$
$$= [1/(\cos x - \varepsilon \sin x)] \cdot [(\cos x + \varepsilon \sin x)/(\cos x + \varepsilon \sin x)]$$
$$= (\cos x + \varepsilon \sin x)/\cos^2 x = \sec x + \varepsilon \sec x \tan x.$$

4. Higher order infinitesimals. Let I be an n-dimensional real vector space with basis $\iota_1, \iota_2, \ldots, \iota_n$. Let $R_n = R + I$, with order, addition, and multiplication defined by:

$$(<) \quad a_0 + a_1\iota_1 + \cdots + a_n\iota_n < b_0 + b_1\iota_1 + \cdots + b_n\iota_n$$

if $a_j < b_j$ at the first subscript j for which $a_j \neq b_j$;

$$(+)\quad (a_0 + a_1\iota_1 + \cdots + a_n\iota_n) + (b_0 + b_1\iota_1 + \cdots + b_n\iota_n)$$
$$= (a_0 + b_0) + (a_1 + b_1)\iota_1 + \cdots + (a_n + b_n)\iota_n;$$
$$(\times)\quad (a_0 + a_1\iota_1 + \cdots + a_n\iota_n) \times (b_0 + b_1\iota_1 + \cdots + b_n\iota_n)$$
$$= (a_0b_0) + (a_0b_1 + a_1b_0)\iota_1 + \cdots + \left(\sum_{i=0}^{n} a_ib_{n-i}\right)\iota_n.$$

Then $\iota_i\iota_j = \iota_{i+j}$ if $i + j \leqslant n$, and $\iota_i\iota_j = 0$ if $i + j > n$, and if $\iota = \iota_1$, a typical element of R_n can be written $a_0 + a_1\iota + a_2\iota^2 + \cdots + a_n\iota^n$, an nth degree polynomial in ι. R_n is a commutative ring with identity, and the ideal $I_k = R_n\iota^k$ is the set of kth *order infinitesimals.* If ε is any element of I, then ε^k is in I_k and $\varepsilon^{n+1} = 0$. Every positive infinitesimal of order $k + 1$ is less than every positive infinitesimal of order k, and they are all less than any positive real number.

R_n has the following properties:

(i) $1/(x + \varepsilon) = (1/x)(1 - \varepsilon/x + \cdots + (-1)^n(\varepsilon^n/x^n))$ for $x \neq 0$;

(ii) $(x + \varepsilon)^r = x^r + rx^{r-1}\varepsilon + \cdots + (r(r - 1) \cdots (r - n + 1)/n!)x^{r-n}\varepsilon^n$ for r rational, $x > 0$.

It follows from (ii) that the natural extension $f_*(x + \varepsilon) = (x + \varepsilon)^r$ of $f(x) = x^r$ to R_n satisfies

$$f_*(x + \varepsilon) = f(x) + f'(x)\varepsilon + \cdots + \left(f^{(n)}(x)/n!\right)\varepsilon^n.$$

DEFINITION. If R is a linearly ordered ring, x in R, then $|x| = x$ if $x \geqslant 0$ and $|x| = -x$ if $x < 0$.

THEOREM 4. *If $Tf = f_*$ is defined on the set A_n of n-times differentiable functions by*

$$f_*(x + \varepsilon) = \sum_{k=0}^{n} \left(f^{(k)}(x)/k!\right)\varepsilon^k,$$

then T is an extension to R_n, and T satisfies the nth order Correspondence Principle, which is that if ϕ is in A_n, and $|\phi(h)| \leqslant M|h|^n$ for all sufficiently small h, then $|\phi_*(\varepsilon)| \leqslant M|\varepsilon|^n$ for all ε.

Proof. T preserves multiplication by scalars, addition, multiplication, and composition, and by $n - 1$ uses of l'Hospital's rule on $\phi(h)/h^n$, $|\phi^{(n)}(0)| \leqslant Mn!$, so $|\phi_*(\varepsilon)| = |(\phi^{(n)}(0)/n!)\varepsilon^n| \leqslant M|\varepsilon|^n$ for all $\varepsilon \neq 0$.

THEOREM 5. *If $Tf = f_*$ is an extension to R_n with domain A_n, which satisfies the nth order Correspondence Principle, and if f is in A_n and f is n-times differentiable at x, then*

$$f_*(x + \varepsilon) = \sum_{k=0}^{n} \left(f^{(k)}(x)/k!\right)\varepsilon^k.$$

Proof. Let M be any positive number. Let

$$\phi_k(h) = f^{(n-k)}(x + h) - \sum_{r=0}^{k} \left(f^{(n-k+r)}(x)/r!\right)h^r.$$

Then

$$|\phi_1(h)| = |f^{(n-1)}(x + h) - f^{(n-1)}(x) - f^{(n)}(\bar{x})h| \leqslant M|h|$$

for all sufficiently small h, and

$$\phi_{k+1}(h) = \int_0^h \phi_k(h')\,dh',$$

so

$$|\phi_{k+1}(h)| \leqslant |h| \max|\phi_k(h')|, \quad -|h| \leqslant h' \leqslant |h|,$$

and so by induction on k,

$$|\phi_n(h)| = \left|f(x+h) - \sum_{r=0}^{n} (f^{(r)}(x)/r!)h^r\right| \leq M|h|^n$$

for all sufficiently small h. Therefore, for all $M > 0$,

$$\left|f_*(x+\varepsilon) - \sum_{r=0}^{n} (f^{(r)}(x)/r!)\varepsilon^r\right| \leq M|\varepsilon|^n, \quad \text{for all } \varepsilon,$$

so

$$f_*(x+\varepsilon) - \sum_{r=0}^{n} (f^{(r)}(x)/r!)\varepsilon^r = 0.$$

Let R_∞ be the set of formal power series $a_0 + a_1\iota + a_2\iota^2 + \cdots$ with the usual addition and multiplication and lexicographic order. Then $R_\infty = R + I$, where $I = R_\infty\iota$. If $Tf = f_*$ is defined on A_∞ by

$$f_*(x+\varepsilon) = f(x) + f'(x)\varepsilon + (f''(x)/2!)\varepsilon^2 + \cdots,$$

then T is an extension to R_∞. Moreover, T satisfies the *Correspondence Principle*: *if ϕ is in A_∞ and $|\phi(h)| \leq M|h|^n$ for all sufficiently small h, then for all ε, $|\phi_*(\varepsilon)| \leq M|\varepsilon|^n + \varepsilon'$, where ε' is of order $n' > n$.*

Proof. Let $M > 0$ and n be given. If $|\phi(h)| \leq M|h|^n$ for all sufficiently small h, then

$$\phi(0) = \phi'(0) = \cdots = \phi^{(n-1)}(0) = 0, \quad \text{and} \quad |\phi^{(n)}(0)| \leq Mn!,$$

so for all ε,

$$|\phi_*(\varepsilon)| \leq \left|\frac{\phi^{(n)}}{n!}(0)\varepsilon^n\right| + \left|\frac{\phi^{(n+1)}}{(n+1)!}(0)\varepsilon^{n+1} + \cdots\right| < M|\varepsilon|^n + \varepsilon'.$$

THEOREM 6. *If $Tf = f_*$ is an extension to R_∞ with domain A_∞, and if T satisfies the Correspondence Principle, then*

$$f_*(x+\varepsilon) = \sum_{n=0}^{\infty} (f^{(n)}(x)/n!)\varepsilon^n.$$

Proof. If f is in A_∞, x in dom(f), then for any n

$$\phi(h) = f(x+h) - \sum_{k=0}^{n} (f^{(k)}(x)/k!)h^k$$

is also in A_∞, and for any $M > 0$, $|\phi(h)| \leq M|h|^n$ for all sufficiently small h, therefore,

$$\phi_*(\varepsilon) = f_*(x+\varepsilon) - \sum_{k=0}^{n} (f^{(k)}(x)/k!)\varepsilon^k \text{ is in } I_n = R_\infty\iota^n$$

and therefore $f_*(x+\varepsilon) - \Sigma_{k=0}^{\infty}(f^{(k)}(x)/k!)\varepsilon^k$ is also in I_n. But the intersection of the I_n, $n = 1,2,\ldots$, is zero, so

$$f_*(x+\varepsilon) = \sum_{k=0}^{\infty} (f^{(k)}(x)/k!)\varepsilon^k.$$

It is possible, but not particularly simple, to compute the higher derivatives of elementary functions algebraically by using Theorem 5. For example, to get the second derivative of $f(x) = (1 + x^3)^{1/2}$ would require at least the computation

$$\begin{aligned}\left[1+(x+\varepsilon)^3\right]^{1/2} &= \left[(1+x^3)+(3x^2\varepsilon+3x\varepsilon^2)\right]^{1/2}\\ &= (1+x^3)^{1/2}+(1/2)(1+x^3)^{-1/2}(3x^2\varepsilon+3x\varepsilon^2)\\ &\quad -(1/8)(1+x^3)^{-3/2}(3x^2\varepsilon+3x\varepsilon^2)^2\\ &= (1+x^3)^{1/2}+(1/2)(1+x^3)^{-1/2}3x^2\varepsilon\\ &\quad +(1/2)\left[(1+x^3)^{-1/2}3x-(9/4)(1+x^3)^{-3/2}x^4\right]\varepsilon^2.\end{aligned}$$

I think the interest of Theorems 5 and 6 lies in the fact that the transformation $f(x) \to \Sigma_{k=0}^{n}(f^{(k)}(x)/k!)\varepsilon^k$ is completely characterized by the Correspondence Principle and the extension properties (preserving scalar multiplication, addition, multiplication, and composition).

THEOREM 7. *If $Tf = f_*$ is any extension to R_n with domain $A_n (n = 1, 2, \ldots, \infty)$ which satisfies the appropriate Correspondence Principle, and if $f' > 0$, then f_* is monotone increasing.*

Proof. Let $\varepsilon = a_1\iota + a_2\iota^2 + \cdots = \varepsilon_1 + \varepsilon_2 + \cdots$. Then

$$\begin{aligned}f(x+\varepsilon) &= f(x)+f'(x)\varepsilon+(f''(x)/2!)\varepsilon^2+\cdots\\ &= f(x)+f'(x)(\varepsilon_1+\varepsilon_2+\cdots)+(f''(x)/2!)(\varepsilon_1+\varepsilon_2+\cdots)^2+\cdots\\ &= f(x)+f'(x)\varepsilon_1+\left(f'(x)\varepsilon_2+(f''(x)/2!)\varepsilon_1^2\right)+\cdots\\ &\quad +\left(f'(x)\varepsilon_n+\text{some function of } x, \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{n-1}\right)+\cdots.\end{aligned}$$

Suppose $x + \varepsilon < x' + \varepsilon'$. Then $x < x'$ or $x = x'$ and $\varepsilon_1 < \varepsilon_1'$ or $x = x'$ and $\varepsilon_1 = \varepsilon_1'$ and $\varepsilon_2 < \varepsilon_2'$ or $\cdots$. In every case, $f(x + \varepsilon) < f(x' + \varepsilon')$.

References

1. James M. Henle and Eugene M. Kleinberg, Infinitesimal Calculus, MIT Press, Cambridge, Mass., and London, 1979.
2. H. Jerome Keisler, Elementary Calculus, Prindle, Weber & Schmidt, Boston, 1976.
3. ______, Foundations of Infinitesimal Calculus, Prindle, Weber & Schmidt, 1976.
4. Abraham Robinson, Non-Standard Analysis, Amsterdam, North Holland, 1966.
5. David Tall, Looking at graphs through infinitesimal microscopes, windows and telescopes, Math. Gaz., 64 (1980) 22–49.
6. Giuseppe Veronese, Fondamenti di geometria, Padova, Tipografia del seminario, 1891.
7. ______, Grundzüge der Geometrie, Leipzig, Teubner, 1894 (translation of [5] by Adolf Schepp).

REMARKS ON SOME PUTNAM PROBLEMS

HARLEY FLANDERS
Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431

When honored with an invitation to review Gleason, Greenwood, and Kelly, *The William Lowell Putnam Mathematical Competition* (GGK) for *The Mathematical Intelligencer*, I took the

---

Harley Flanders did his doctoral work at the University of Chicago under André Weil and Otto Schilling. He has worked in class field theory, differential geometry, linear algebra, electric network theory, and computing. He has held positions at Cal Tech, U.C. Berkeley, Purdue, and Tel Aviv University and visiting positions at Cambridge University, University of Washington, University College London, University of Wisconsin, Georgia Tech, and Florida Atlantic University. His MAA activities include chairing two Sections: N. California and Indiana, and editing this MONTHLY. Hobbies include music, chess, ping-pong, woodworking, and sailing.

opportunity to read all of the solutions rather carefully and to compare them with my own notes, accumulated over many years. In some cases my approach is sufficiently different from that of GGK that it may be of interest.

1. Differential Equations.

PUTNAM COMPETITION #13 (1953), PROBLEM PM3 (GGK, p. 375).

Solve the equations

$$\frac{dy}{dx} = z(y+z)^n \qquad \frac{dz}{dx} = y(y+z)^n,$$

given the initial conditions $y = 1$ and $z = 0$ when $x = 0$.

GGK uses $d(y+z)/dx = (y+z)^{n+1}$, to obtain the integral $(y+z)^{-n} = a - nx$ (or $y + z = ae^x$ when $n = 0$), but overlooks $y\,dy/dx = z\,dz/dx$, yielding $y^2 = z^2 + b$. The solution follows easily.

PUTNAM COMPETITION #20 (1959), PROBLEM AM5 (GGK, p. 500).

A sparrow, flying horizontally in a straight line, is 50 feet directly below an eagle and 100 feet directly above a hawk. Both hawk and eagle fly directly toward the sparrow, reaching it simultaneously. The hawk flies twice as fast as the sparrow. How far does each bird fly? At what rate does the eagle fly?

The solution in GGK presupposes that each bird flies at a uniform speed. Actually, it suffices to assume that the eagle's speed is a constant times that of the sparrow.

Suppose the sparrow starts at (0, 0) and moves forward along the x-axis, so his position at time t is $(u, 0)$. Let the pursuer start at $(0, a)$, with $a > 0$, and speed k times that of the sparrow, where $k > 1$. Then

$$\begin{cases} dx^2 + dy^2 = k^2 du^2 \\ y\,dx + (u-x)\,dy = 0 \end{cases} \qquad \begin{cases} \text{initial:} & x = u = 0,\ y = a \\ \text{terminal:} & x = u = b,\ y = 0. \end{cases}$$

Clearly $u - x > 0$, $dx/du > 0$, $dy/du < 0$ for $0 < u < b$. Appropriate polar coordinates expedite the integration: set $dx/du = k\cos\theta$, $dy/du = -k\sin\theta$; then set $y = r\sin\theta$, $u - x = r\cos\theta$. One finds that r decreases from a to 0 and θ decreases from $\frac{1}{2}\pi$ to 0. By differentiating the relations defining r and solving for dr/du and $d\theta/du$, we obtain

$$\begin{cases} dr/du = -k + \cos\theta \\ r\,d\theta/du = -\sin\theta, \end{cases}$$

hence

$$\frac{dr}{r} = \frac{k - \cos\theta}{\sin\theta}\,d\theta,$$

which integrates to $r\sin\theta = a(\csc\theta - \cot\theta)^k$. Hence

$$y = a(\csc\theta - \cot\theta)^k \text{ and } \frac{dy}{du} = -k\sin\theta.$$

We have

$$-kb = -k\int_0^b du = \int_{u=0}^{u=b}\frac{dy}{\sin\theta} = \frac{y}{\sin\theta}\bigg|_{u=0}^{u=b} + \int_{u=0}^{u=b}\frac{y\cos\theta}{\sin^2\theta}\,d\theta,$$

$$kb = a + a\int_0^{\pi/2}\frac{\cos\theta(\csc\theta - \cot\theta)^k}{\sin^2\theta}\,d\theta.$$

By the half-angle substitution, $w = \tan\frac{1}{2}\theta$,

$$kb = a + a\int_0^1 \frac{1}{2}(1 - t^2)t^{k-2}\,dt = a + \frac{a}{k^2 - 1} = \frac{ak^2}{k^2 - 1}.$$

Hence

$$b = \frac{ak}{k^2 - 1}$$

from which everything follows routinely.

2. Integration.

PUTNAM COMPETITION #8 (1948), PROBLEM AM4 (GGK, p. 247).

Let D be a plane region bounded by a circle of radius r. Let (x, y) be a point of D and consider a circle of radius δ and center at (x, y). Denote by $l(x, y)$ the length of that arc of the circle which is outside D. Find

$$\lim_{\delta\to 0} \frac{1}{\delta^2} \iint_D l(x, y)\,dx\,dy.$$

A heuristic solution is worth noting. For δ very small relative to r, replace the fixed circle by a straight segment of length $2\pi r$. The required limit is

$$\begin{aligned} A &= \lim_{\delta\to 0} \frac{2\pi r}{\delta^2}\left[2\delta\int_0^\delta \arccos\left(\frac{\delta - x}{\delta}\right) dx\right] \\ &= \lim_{\delta\to 0}\left(4\pi r\int_0^1 \arccos(1 - t)\,dt\right) = 4\pi r\int_0^1 \arccos t\,dt. \end{aligned}$$

By a glance at the cosine graph,

$$\int_0^1 \arccos t\,dt = \int_0^{\pi/2} \cos\theta\,d\theta = 1,$$

so $A = 4\pi r$. (This can be made rigorous of course.)

PUTNAM COMPETITION #17 (1957), PROBLEM PM3 (GGK, p. 446).

For $f(x)$ a positive, monotone decreasing function defined in $0 \leqslant x \leqslant 1$ prove that

$$\frac{\int_0^1 xf^2(x)\,dx}{\int_0^1 xf(x)\,dx} \leqslant \frac{\int_0^1 f^2(x)\,dx}{\int_0^1 f(x)\,dx}.$$

GGK gives a beautiful solution based on integrating $f(x)f(y)(y - x)[f(x) - f(y)]$ over the unit square. A related one variable proof is possible starting with

$$F(x) = \int_0^x f^2(t)\,dt\int_0^x tf(t)\,dt - \int_0^x tf^2(t)\,dt\int_0^x f(t)\,dt.$$

One finds $F(0) = 0$ and

$$F'(x) = f(x)\int_0^x (t - x)[f(x) - f(t)]f(t)\,dt \geqslant 0,$$

so $F(1) \geqslant 0$.

3. Limits.

PUTNAM COMPETITION #13 (1953), PROBLEM AM6 (GGK, p. 370).

Show that the sequence

$$\sqrt{7}, \sqrt{7-\sqrt{7}}, \sqrt{7-\sqrt{7+\sqrt{7}}}, \sqrt{7-\sqrt{7+\sqrt{7-\sqrt{7}}}}, \ldots$$

converges and evaluate the limit.

GGK proves $f(x)=\sqrt{7-\sqrt{7+x}}$ contractive by using the mean value theorem. Another possibility is to use the system

$$\begin{cases} x_{n+1}=\sqrt{7-y_n} \\ y_{n+1}=\sqrt{7+x_n} \end{cases} \qquad \begin{cases} x_0=0 \\ y_0=0. \end{cases}$$

By induction one easily shows $0 \leqslant x_n < 7, 0 \leqslant y_n < 7$. From $x_{n+1}^2 = 7 - y_n$ and $2^2 = 7-3$ follows $x_{n+1}^2 - 2^2 = 3 - y_n$, so

$$|x_{n+1}-2| = \frac{|y_n-3|}{x_{n+1}+2} \leqslant \frac{1}{2}|y_n-3|.$$

Similarly $|y_{n+1}-3| \leqslant \frac{1}{3}|x_n-2|$. It follows readily that $x_n \to 2$ (and $y_n \to 3$).

PUTNAM COMPETITION #19 (1958), PROBLEM PM4 (GGK, p. 490).

Let C be a real number, and let f be a function such that

$$\lim_{x\to\infty} f(x) = C, \quad \lim_{x\to\infty} f'''(x) = 0.$$

Prove that

$$\lim_{x\to\infty} f'(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} f''(x) = 0,$$

where superscripts denote derivatives.

GGK uses Taylor's theorem, and it is not obvious how to extend the proof to show (J. E. Littlewood) that if $n \geqslant 1$, if $f(x) \to C$, and if $f^{(n+1)}(x)$ is bounded as $x \to \infty$, then $f^{(n)}(x) \to 0$. A systematic approach is to use the operators

$$Ef(x) = f(x+h), \quad \Delta = E - 1.$$

Then

$$(E-1)^n f = \Delta^n f = h^n f^{(n)}(x+\xi),$$

where $0 < \xi < nh$. Clearly $(E-1)^n f \to 0$ as $x \to \infty$, so each interval $[x, x+nh]$ with x sufficiently large contains a point where $|f^{(n)}| < \xi$. Choosing h sufficiently small then guarantees $|f^{(n)}| < 2\varepsilon$ on this whole interval, by one application of the MVT plus the boundedness of $f^{(n+1)}$, etc.

4. Inequalities.

PUTNAM COMPETITION #25 (1964), PROBLEM AM5 (GGK, p. 589).

Prove that there is a positive constant K such that the following inequality holds for any sequence of positive numbers $a_1, a_2, a_3, \ldots$:

$$\sum_{n=1}^{\infty} \frac{n}{a_1+a_2+\cdots+a_n} \leqslant K \sum_{n=1}^{\infty} \frac{1}{a_n}.$$

K. Knopp proved this with $K = 4$. GGK gives two interesting solutions, the second due to R. M. Redheffer. The following is based on the Cauchy-Schwarz inequality. Let $\mathbf{u} = (\sqrt{a_1}, \dots, \sqrt{a_n})$ and $\mathbf{v} = (1/\sqrt{a_1}, 2/\sqrt{a_2}, \dots, n/\sqrt{a_n})$. Then $(\mathbf{u} \cdot \mathbf{v})^2 \leq |\mathbf{u}|^2 |\mathbf{v}|^2$, so

$$(1 + 2 + \cdots + n)^2 \leq (a_1 + \cdots + a_n)\left(\frac{1^2}{a_1} + \frac{2^2}{a_2} + \cdots + \frac{n^2}{a_n}\right).$$

Hence

$$\frac{2n+1}{a_1 + \cdots + a_n} \leq 4\frac{2n+1}{n^2(n+1)^2} \sum_{j=1}^{n} \frac{j^2}{a_j},$$

$$\sum_{n=1}^{\infty} \frac{2n+1}{a_1 + \cdots + a_n} \leq 4 \sum_{n=1}^{\infty} \frac{2n+1}{n^2(n+1)^2} \sum_{j=1}^{n} \frac{j^2}{a_j}$$

$$= 4 \sum_{j=1}^{\infty} \frac{j^2}{a_j} \sum_{n=j}^{\infty} \frac{2n+1}{n^2(n+1)^2}.$$

But

$$\sum_{n=j}^{\infty} \frac{2n+1}{n^2(n+1)^2} = \sum_{n=j}^{\infty} \left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right) = \frac{1}{j^2},$$

hence

$$\sum_{n=1}^{\infty} \frac{2n+1}{a_1 + \cdots + a_n} \leq 4 \sum_{j=1}^{\infty} \frac{1}{a_j}$$

(Redheffer). The required inequality, with $K = 2$, follows.

PUTNAM COMPETITION #21 (1960), PROBLEM AM3 (GGK, p. 518).

Show that if t_1, t_2, t_3, t_4, t_5 are real numbers, then

$$\sum_{j=1}^{5} (1 - t_j) \exp\left(\sum_{k=1}^{j} t_k\right) \leq e^{e^{e^e}}.$$

More generally, define $e_0 = 0$ and $e_n = \exp(e_{n-1})$. Then for $n \geq 1$, we shall prove

$$\sum_{j=1}^{n} (1 - t_j) \exp\left(\sum_{k=1}^{j} t_k\right) \leq e_n$$

with equality if and only if $t_n = e_0, t_{n-1} = e_1, \dots, t_1 = e_{n-1}$.

For $n = 1$, the inequality $(1 - t)e^t \leq 1$ is equivalent to $1 - t \leq e^{-t}$, that is, to $1 + t \leq e^t$, which is obvious from the convex graph of $y = e^x$. Equality if and only if $t = 0 = e_0$. If the inequality is true for an $n - 1 \geq 0$, then

$$\sum_{j=1}^{n} (1 - t_j) \exp\left(\sum_{k=1}^{j} t_k\right) = \exp(t_1)\left[(1 - t_1) + \sum_{j=2}^{n} (1 - t_j) \exp\left(\sum_{k=2}^{j} t_k\right)\right]$$

$$\leq \exp(t_1)\left[(1 - t_1) + e_{n-1}\right],$$

with equality if and only if $t_n = e_0, \dots, t_2 = e_{n-2}$. But

$$\exp(t_1)\left[(1 - t_1) + e_{n-1}\right] = \exp(t_1)\left[1 - (t_1 - e_{n-1})\right]$$
$$\leq \exp(t_1)\exp(e_{n-1} - t_1) = \exp e_{n-1} = e_n,$$

with equality if and only if $t_1 = e_{n-1}$.

5. A Trigonometric Polynomial.

PUTNAM COMPETITION #23 (1962), PROBLEM PM6 (GGK, p. 567).

Let

$$f(x) = \sum_{k=0}^{n} a_k \sin kx + b_k \cos kx,$$

where a_k and b_k are constants. Show that if $|f(x)| \leqslant 1$ for $0 \leqslant x \leqslant 2\pi$ and $|f(x_j)| = 1$ for $0 \leqslant x_1 < x_2 < \cdots < x_{2n} < 2\pi$, then $f(x) = \cos(nx + a)$ for some constant a.

GGK uses a differential equation. Instead, start with the fact that an nth degree trigonometric polynomial has at most $2n$ zeros. Clearly $f'(x_j) = 0$, so $f'(x)$ has no further zeros, hence $f(x)$ is monotone between its maxs and mins. This implies

$$f(x_1) = f(x_3) = \cdots = f(x_{2n-1}) = +1$$
$$f(x_2) = f(x_4) = \cdots = f(x_{2n}) = -1$$

(or the opposite signs). Set $g(x) = f(x) - f(x + 2\pi/n)$. Then $g(x_{2j-1}) \geqslant 0$ and $g(x_{2j}) \leqslant 0$, so $g(x)$, a trigonometric polynomial of degree at most $n - 1$, has a zero on each interval

$$[x_1, x_2], \ldots, [x_{2n-1}, x_{2n}], [x_{2n}, x_1 + 2\pi].$$

A zero at x_j implies $f(x_j) = f(x_j + 2\pi/n) = \pm 1$, so $f'(x_j) = f'(x_j + 2\pi/n) = 0$ (max or min of f), hence $g'(x_j) = 0$, that is, x_j is a double zero or worse of $g(x)$. Thus, counting multiplicities, $g(x)$ has at least $2n$ zeros on $[0, 2\pi]$, hence $g(x) = 0$, so $f(x + 2\pi/n) = f(x)$. This implies

$$a_k \sin(x + 2\pi/n) + b_k \cos(x + 2\pi/n) = a_k \sin kx + b_k \cos kx$$

for $0 < k < n - 1$, which easily forces $a_k = b_k = 0$. The rest is routine.

6. A Combinatorial Identity.

PUTNAM COMPETITION §23 (1962), PROBLEM AM5 (GGK, p. 560).

Evaluate in closed form

$$\sum_{k=1}^{n} \binom{n}{k} k^2.$$

In addition to the two analytic solutions GGK gives, a direct counting solution is possible: The expression represents the number of ways to choose a committee, its chairman, and its secretary (possibly the same) from n people. You can choose chairman = secretary in n ways and the rest of the committee in 2^{n-1} ways, or chairman $\neq$ secretary in $n(n-1)$ ways and the rest in 2^{n-2} ways, so the required sum equals

$$n \cdot 2^{n-1} + n(n-1) \cdot 2^{n-2} = n(n+1) \cdot 2^{n-2}.$$

References

1. A. M. Gleason, R. E. Greenwood, and L. M. Kelly, The William Lowell Putnam Mathematical Competition —Problems and Solutions: 1938-1964, Math. Assoc. of Amer., 1980.

2. H. Flanders, review of the above book, Math. Intelligencer, to appear.

Two of the founders of modern analysis. Their names are on page 505.

$$y = y + y^r(x-y)^s \Rightarrow y^r(x-y)^s = 0 \Rightarrow x = y. \tag{6}$$

In the general case a closed form expression for x does not seem to be readily available. If $r = 1$, however, we may express x as an infinite product. Define

$$d_n = x_n - y_n,\ d = d_0 = x_0 - y_0 = t_2 - t_1. \tag{7}$$

Then, subtracting (3) from (2), we obtain:

$$d_{n+1} = d_n^{s+1}. \tag{8}$$

It follows by an easy induction that

$$d_n = d^{(s+1)^n}. \tag{9}$$

We note that

$$y_{n+1}/y_n = 1 + d_n^s. \tag{10}$$

Forming products on both sides of (10) by letting n approach infinity, we obtain the limiting expression (which we know exists):

$$y = y_0 \prod_{n=0}^{\infty} \{1 + d^{s(s+1)^n}\}. \tag{11}$$

In the particular case $r = s = 1$, the expression in (11) may be evaluated in closed form, for then

$$y = y_0 \prod_{n=0}^{\infty} (1 + d^{2^n}) = y_0 \prod_{n=0}^{\infty} \left\{ \frac{1 - d^{2^{n+1}}}{1 - d^{2^n}} \right\} = y_0(1-d)^{-1}, \quad \text{or}$$

$$y = y_0(1 - x_0 + y_0)^{-1}. \tag{12}$$

Also solved by L. Kuipers (Switzerland) and Michael Skalsky.

A Set of Homogeneous Equations

6312* [1980, 675]. *Proposed by M. S. Klamkin, University of Alberta.*

Prove or disprove that the set of n equations in n unknowns

$$x_1^{l_i} + x_2^{l_i} + \cdots + x_n^{l_i} = 0 \qquad (i = 1, 2, \ldots, n),$$

where the l_i's are relatively prime positive integers, has only the trivial solution $x_i = 0$ $(i = 1, 2, \ldots, n)$, if and only if each $m = 2, 3, \ldots, n$ divides at least one l_i.

Solution by Constantine Nakassis, Gaithersburg, Maryland. Let $n > 2$ be an even number $(n = 2k)$; suppose that the only even number in $l_1, l_2, \ldots, l_n$ is l_1 (for example, take $l_1 = n!$ and let $l_2, \ldots, l_n$ be the first $(n-1)$ primes that follow n). Consider any k complex numbers which satisfy

$$y_1^{l_1} + y_2^{l_1} + \cdots + y_k^{l_1} = 0.$$

Let $x_{2i-1} = y_i$, $x_{2i} = -y_i$ for $i = 1, 2, \ldots k$. It is clear then that the proposed system has nontrivial solutions. (The starred assertion is true if $n = 2$, but false if $n = 2k + 1 > 3$.)

The case $n = 3$ remains open; the starred assertion has been established by the proposer for many triples.

ANSWERS TO "PHOTOS" ON PAGE 455

Top: Israil Moiseevich Gelfand; bottom: Mark Aronovich Naimark.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

AT WHAT POINTS IS THE PROJECTION MAPPING DIFFERENTIABLE?

J.-B. HIRIART-URRUTY

U. E. R. Mathématiques, Informatique, Gestion, Université Paul Sabatier, (Toulouse III) 118, Route de Narbonne, 31062-Toulouse Cedex, France

1. Given a nonempty closed convex set Q of $\mathbb{R}^n$, the projection mapping P_Q is defined as assigning to each $x \in \mathbb{R}^n$ the closest point to x in Q, so that

$$\|x - P_Q(x)\| = \operatorname{Min}\{\|x - y\|, y \in Q\}. \tag{1}$$

Many papers have been devoted to the properties of P_Q in recent years; a thorough account of the known facts is given in the long paper by Zarantonello [**12**]. A fundamental property of P_Q is that it is **nonexpansive**, that is

$$\|P_Q(x) - P_Q(x')\| \leqslant \|x - x'\|, \qquad x, x' \in \mathbb{R}^n. \tag{2}$$

According to Rademacher's theorem (1919), this Lipschitz property guarantees that P_Q is differentiable almost everywhere (a.e.), i.e., except on a set of Lebesgue measure zero. Therefore, a question which arises quite naturally is:

A1. *At what points is P_Q differentiable?*

In spite of the elementary formulation of this question, a full answer is so far unknown. To get a better insight into the kind of problems which come up, we shall briefly indicate what is known on this subject. To begin with, we specify that only the finite-dimensional case is considered here and that the differentiability of P_Q should be understood in the Fréchet sense. It entails no loss of generality to assume that Q has a nonempty interior and is different from the whole space $\mathbb{R}^n$. If $x \in \operatorname{int} Q$, clearly $P_Q(x') = x'$ in a neighborhood of x so that P_Q is differentiable at x and $DP_Q(x)$ is the identity map. If $\bar{x} \in \operatorname{bd} Q$, it can be shown [**9**], [**12**] that

$$P_Q(\bar{x} + th) = \bar{x} + tP_{T(Q,\bar{x})}(h) + o(t), \qquad t > 0, h \in \mathbb{R}^n,$$

where $T(Q, \bar{x})$ is the **tangent cone** to Q at $\bar{x}$, i.e.,

$$T(Q, \bar{x}) = \operatorname{cl} \bigcup_{\lambda > 0} \lambda(Q - \bar{x}).$$

Thus, at each point $\bar{x} \in \operatorname{bd} Q$, the map P_Q has a directional derivative in all directions h, given by $P_{T(Q,\bar{x})}(h)$. But, under the assumptions made on Q, $T(Q, \bar{x})$ cannot be a linear space and P_Q is not differentiable at $\bar{x}$. Note that $\operatorname{bd} Q$ is exactly the set (of measure zero) where the distance function d_Q is not differentiable. The question A1 can therefore be rephrased as follows:

At what points $x \notin Q$ is P_Q differentiable?

Actually, in consideration of the differentiability property of P_Q, the set $U = \mathbb{R}^n \setminus Q$ can be partitioned in open half-lines as follows: if P_Q is differentiable at $x \in U$, then P_Q is differentiable at each point x' of the form

$$x' = \lambda x + (1 - \lambda)P_Q(x), \qquad \lambda > 0,$$

that is, for each point x' on the open half-line from $P_Q(x)$ through x [4], [7]. We recall that if $\bar{x} \in \operatorname{bd} Q$, the **normal cone** $N(Q, \bar{x})$ to Q at $\bar{x}$ consists of all x^* for which the linear form $\langle x^*, . \rangle$ attains its maximum value over Q at $\bar{x}$, i.e.,

$$N(Q, \bar{x}) = T(Q, \bar{x})^0 = \{x^* \mid \langle x^*, x - \bar{x} \rangle \leq 0 \text{ for all } x \in Q\}.$$

The collection $\{\bar{x} + N(Q, \bar{x}) \mid \bar{x} \in \operatorname{bd} Q\}$ is a collection of closed convex cones having the following properties [12]:

(i) the collection is pairwise disjoint,
(ii) the union of the whole collection is the closure of U.

Moreover, $\bar{x} + N(Q, \bar{x})$ is the preimage of $\{\bar{x}\}$ under the mapping P_Q. If $x \in U$ lies in the interior of $\bar{x} + N(Q, \bar{x})$ for some $\bar{x} \in \operatorname{bd} Q$, it is clear that $P_Q(x') = \bar{x}$ for every x' in a neighborhood of x and thus $DP_Q(x) = 0$. On the other hand, if x lies on the boundary of $\bar{x} + N(Q, \bar{x})$, we have seen that the differentiability or nondifferentiability of P_Q at x is propagated to the whole open half-line $\{\lambda x + (1 - \lambda)\bar{x} \mid \lambda > 0\}$. At a point where P_Q is not differentiable, it is not even certain that it admits a directional derivative in all directions [8]. The question we have posed is thus *to detect the points in* $\operatorname{bd}(\bar{x} + N(Q, \bar{x}))$ *where* P_Q *is differentiable, for each* $\bar{x} \in \operatorname{bd} Q$. People have mainly focused their attention on the necessary (and sufficient) conditions on $\operatorname{bd} Q$ for P_Q to be continuously differentiable (C^1) on U and the resulting expression of $DP_Q(x)$ [4], [7]. Simple examples and drawings will convince the reader that the differentiability of P_Q is strongly related to the smoothness of the boundary of Q. An appropriate definition of "C^2 boundary" is given in terms of gauge function of Q. Recall that if $x_0 \in \operatorname{int} Q$, then the **gauge function** μ_Q for Q with respect to x_0 is defined by

$$\mu_Q(x) = \inf\{t > 0 \mid x - x_0 \in t(Q - x_0)\}, \qquad x \in \mathbb{R}^n.$$

Now, $\operatorname{bd} Q$ is said to be C^2 if for some $x_0 \in \operatorname{int} Q$ the gauge function μ_Q with respect to x_0 is twice continuously differentiable in some neighborhood of $\operatorname{bd} Q$. This definition, which is used in [4] and [7], turns out to be independent of the choice of x_0 and, moreover, Holmes [7] has shown that it is equivalent to $\operatorname{bd} Q$ being locally a C^2-submanifold of $\mathbb{R}^n$, appropriately modelled on some hyperplane in $\mathbb{R}^n$. The next sufficient condition for P_Q to be C^1 on U has been proved by Holmes [7].

THEOREM 1. *If Q has a C^2 boundary, then P_Q is C^1 on U.*

Holmes also obtained a formula for $DP_Q(x)$ and, in doing so, proved the invertibility of the restriction of $DP_Q(x)$ to the hyperplane Π_x orthogonal to the (nonnull) vector $x - P_Q(x)$, that is

$$\Pi_x = \{x \in \mathbb{R}^n \mid \langle x^*, x - P_Q(x) \rangle = 0\}.$$

In their pithy paper [4], Fitzpatrick and Phelps used this latter property to prove the following converse to Theorem 1:

THEOREM 2. *Suppose that P_Q is C^1 on U, with the restriction of $DP_Q(x)$ to Π_x invertible for each $x \in U$; then the boundary of Q is C^2.*

Thus, to ensure that P_Q is C^1 in a neighborhood of the open ray normal to Q at $\bar{x} \in \operatorname{bd} Q$, it is almost necessary and sufficient to assume that $\operatorname{bd} Q$ is C^2 near the point $\bar{x}$. But, when $\bar{x}$ is some "corner point" of $\operatorname{bd} Q$, the classification of the open normal rays emanating from $\bar{x}$ in terms of differentiability properties of P_Q remains to be done.

2. If one knew at what points P_Q is differentiable and the expression of $DP_Q(x)$ at these points, one would be able to answer a related question concerning the generalized derivative of P_Q in the sense of Clarke [3]. Let us recall that the **generalized derivative** of P_Q at $x \in \mathbb{R}^n$, denoted by $\partial P_Q(x)$, is the set of linear mappings defined as the convex hull of the set

$$\{L \mid \exists x_i \to x \text{ with } P_Q \text{ differentiable at } x_i \text{ and } DP_Q(x_i) \to L\}.$$

As is readily seen, this definition yields a nonempty convex compact set of linear mappings which reduces to $\{DP_Q(x)\}$ whenever P_Q is C^1 at x. The charm of this generalized derivative lies in the fact that it is defined at each $x \in \mathbb{R}^n$. A convenient way to handle $\partial P_Q(x)$ or, more precisely, the collection of images $\{\partial P_Q(x)u \mid u \in \mathbb{R}^n\}$ is to use the following characterization of those images [**6**]:

$$\max\{\langle Lu, v\rangle \mid L \in \partial P_Q(x)\} = \limsup_{\substack{x' \to x \\ t \to 0^+}} \frac{\langle P_Q(x'+tu) - P_Q(x'), v\rangle}{t} \tag{3}$$

for all $u, v \in \mathbb{R}^n$.

When $x \in \operatorname{int} Q$, we already know that $\partial P_Q(x)$ reduces to a singleton, containing the identity mapping. When $x \notin \operatorname{int} Q$, examples suggest that the "size" or "dimension" of $\partial P_Q(x)$ (or of the sets $\partial P_Q(x)u$) would give useful information on the structure of bd Q near the point $P_Q(x)$. Some properties of $\partial P_Q(x)$ can readily be derived from those known about $DP_Q(x_i)$ for a sequence (x_i) such as appears in the definition; nevertheless the following question is unsolved:

A2. *What is the structure of the generalized derivative of P_Q at $x \notin$ int Q?*

3. When Q is merely *closed*, the set $P_Q(x)$ of closest points to x in Q is included in the subdifferential at x of an appropriate convex function (see [**2**] and references therein). The projection multifunction P_Q is thus "differentiable" a.e. [**10**, §1.2]. Whence the question A1 for nonconvex Q.

References

1. T. J. Abatzoglou, The minimum norm projection of C^2-manifolds in R^n, Trans. Amer. Math. Soc., 243 (1978) 115–122.

2. E. Asplund, Differentiability of the metric projection in finite-dimensional Euclidean space, Proc. Amer. Math. Soc., 38 (1973) 218–219.

3. F. H. Clarke, On the inverse function theorem, Pacific J. Math., 64 (1976) 97–102.

4. S. Fitzpatrick and R. R. Phelps, Differentiability of the metric projection in Hilbert space, to appear.

5. A. Haraux, How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities, J. Math. Soc. Japan, 29 (1977) 615–631.

6. J.-B. Hiriart-Urruty, Characterizations of the plenary hull of the generalized Jacobian matrix, Math. Programming, Study 18 (D. Sorensen and R. J.-B. Wets, editors), to appear in 1982.

7. R. B. Holmes, Smoothness of certain metric projections on Hilbert space, Trans. Amer. Math. Soc., 184 (1973) 87–100.

8. J. Kruskal, Two convex counterexamples: a discontinuous envelope function and a nondifferentiable nearest point mapping, Proc. Amer. Math. Soc., 23 (1969) 697–703.

9. G. P. McCormick and R. Tapia, The gradient projection method under mild differentiability conditions, S.I.A.M. J. Control, 10 (1972) 93–98.

10. F. Mignot, Contrôle dans les inèquations variationnelles elliptiques, J. Functional Analysis, 22 (1976) 130–185.

11. J.-J. Moreau, Proximitè et dualitè dans un espace Hilbertien, Bull. Soc. Math. France, 93 (1965) 273–299.

12. E. H. Zarantonello, Projections on convex sets in Hilbert space and spectral theory, Contributions to Nonlinear Functional Analysis, Publn. no. 27, Math. Res. Center, University of Wisconsin, Academic Press (1971) 237–424.

THE OPEN MAPPING THEOREM

JOCHEM ZOWE
Fakultät für Mathematik und Physik der Universität Bayreuth, Universitätsstrasse 30, D-8580 Bayreuth, West Germany

Let T be a continuous linear operator between Banach spaces X and Y and X_ρ, Y_ρ closed balls

around zero in X and Y, respectively, with radius $\rho > 0$. Then the open mapping theorem states that $TX = \{Tx \mid x \in X_\rho\}$ covers Y_1 for some $\rho > 0$, whenever T is surjective:

$$TX = Y \Rightarrow \exists \rho > 0 \colon TX_\rho \supset Y_1. \tag{1}$$

Now, in addition, let a closed convex cone K in Y with vertex at zero be given. Then the following generalization of (1) can be shown (cf. **[6]**; see also **[1]**, **[2]**, **[4]** and **[5]** Theorem 1):

If

$$TX + K = Y, \tag{2}$$

then there is $\rho > 0$ such that

$$TX_\rho + K_\rho \supset Y_1; \tag{3}$$

here $TX + K := \{Tx + k \mid x \in X,\, k \in K\}$ and $K_\rho := K \cap Y_\rho$.

Examples satisfying (2) are

$$TX = Y \text{ and } K \text{ arbitrary} \tag{4}$$

or

$$\operatorname{int} K \neq \varnothing \text{ and } T\bar{x} \in -\operatorname{int} K \text{ for some } \bar{x} \in X \tag{5}$$

or, if Y is the product of two subspaces, then a combination of (4) and (5): (4) holds for one subspace and (5) for the other; see condition (6) below. But, the implication "(5) $\Rightarrow$ (3)" is trivial, "(4) $\Rightarrow$ (3)" is identical with (1) and the implication "(6) $\Rightarrow$ (3)" follows from (1) already, together with some simple calculation. Hence, to verify that "(2) $\Rightarrow$ (3)" is a proper strengthening of (1) we are led to the

Problem: *Give an example where* (2) *holds but not* (6), *or show that* (2) *cannot be satisfied unless the situation* (6) *is given.*

For the formulation of condition (6) we write $Y = Y^{(1)} \times Y^{(2)}$ if Y is the topological product of its subspaces $Y^{(1)}, Y^{(2)}$; further, let $T = (T^{(1)}, T^{(2)})$ denote the corresponding decomposition of T and put $K^{(1)} := K \cap Y^{(1)}$, $K^{(2)} := K \cap Y^{(2)}$. We say (6) holds whenever

$$\begin{aligned}&Y \text{ has a decomposition } Y = Y^{(1)} \times Y^{(2)} \text{ such that } \operatorname{int}_{Y^{(1)}} K^{(1)} \neq \varnothing,\, K^{(2)} \text{ is arbitrary,}\\ &T^{(2)}X = Y^{(2)} \text{ and there is } \bar{x} \in X \text{ with } T^{(1)}\bar{x} \in -\operatorname{int}_{Y^{(1)}} K^{(1)},\ T^{(2)}\bar{x} = 0_{Y^{(2)}}.\end{aligned} \tag{6}$$

(4) and (5) are special cases of (6). It is easy to see that (6) implies (2) and thus (3). But, as mentioned above, the implication "(6) $\Rightarrow$ (3)" follows from (1) already. Choose $\alpha > 0$ and $\beta > 0$ such that

$$T^{(2)}X_\alpha \supset Y_1^{(2)} \quad 0_{Y^{(1)}} \in \left(T^{(1)}\bar{x} + K_\beta^{(1)}\right) \text{ where } K_\beta^{(1)} = K^{(1)} \cap Y_\beta^{(1)}.$$

Then with some $\gamma > 0$

$$Y_1^{(1)} - T^{(1)}X_\alpha \subset \gamma\left(T^{(1)}\bar{x} + K_\beta^{(1)}\right).$$

It follows that for $(y^{(1)}, y^{(2)}) \in Y_1^{(1)} \times Y_1^{(2)}$ there are $x \in X_\alpha$ and $k \in K_{\beta\gamma}^{(1)}$ such that

$$y^{(2)} = T^{(2)}x,\ y^{(1)} - T^{(1)}x = T^{(1)}\gamma\bar{x} + k,$$

i.e.,

$$\left(y^{(1)}, y^{(2)}\right) = \left(T^{(1)}(\gamma\bar{x} + x), T^{(2)}(\gamma\bar{x} + x)\right) + (k, 0).$$

If we choose $\rho \geqslant \gamma\beta$ such that $\gamma\bar{x} + X_\alpha \subset X_\rho$, then

$$Y_1 = Y_1^{(1)} \times Y_1^{(2)} \subset (T^{(1)}, T^{(2)})X_\rho + K_\rho.$$

We mention two important possible applications of the above "generalization" of the open mapping theorem. In the context of mathematical programming, (6) corresponds to a so-called

constraint qualification. The implication "(6) $\Rightarrow$ (2)" is used to guarantee certain stability properties of the underlying problem (cf., e.g., [1], [5], [6]). If there are nonpathological situations where (2) holds but not (6), then (2) could be considered as a proper weakening of the well-known constraint qualifications.

The second application has to do with the extension of positive linear functionals. Suppose M is a linear subspace of Y, int $K \neq \varnothing$ and $M \cap$ int $K \neq \varnothing$. Then, as is well known in the standard literature (see, e.g., [3]), a continuous linear functional defined on M and nonnegative on $M \cap K$ has a continuous linear extension to Y which is nonnegative on K. In [6] it is shown that such a continuous positive extension is also possible if the assumption "int $K \neq \varnothing$ and $M \cap$ int $K \neq \varnothing$" is replaced by "$M + K = Y$". The proof uses in a decisive way the implication (2) $\Rightarrow$ (3) to guarantee the continuity of the extension (here M plays the role of X and T is the injection of M into Y). Again, one is interested in an example where $M + K = Y$ but neither $M = Y$ nor $M \cap$ int $K \neq \varnothing$ (or any other trivial situation described by (6)).

References

1. W. Alt, Stabilität mengenwertiger Abbildungen mit Anwendungen auf nichtlineare Optimierungsprobleme, Bayreuther Mathematische Schriften, 3 (1979).
2. P. C. Duong and H. Tuy, Stability, surjectivity and local invertibility of nondifferentiable mappings, Acta Math. Vietnam., 3 (1978) 89–105.
3. G. Jameson, Ordered linear spaces, Lecture Notes in Mathematics 104, Springer, Berlin-Heidelberg-New York, 1972.
4. P. Michel, Problème des inegalités. Applications à la programmation et contrôle optimal, Bull. Soc. Math. France, 101 (1973) 413–439.
5. S. M. Robinson, Regularity and stability for convex mutlivalued functions, Math. of Oper. Research, 1 (1976) 130–143.
6. J. Zowe and S. Kurcyusz, Regularity and stability for the mathematical programming problem in Banach spaces, Appl. Math. Optim., 5 (1979) 49–62.

TWO PROBLEMS IN THE PLANE

P. R. Scott
Department of Pure Mathematics, University of Adelaide, Adelaide, South Australia 5000

Let Λ denote the lattice of integer points in the plane, and let K be a convex set which contains the origin O, but no nonzero points of Λ, in its interior.

A classical theorem of Minkowski states that if K is symmetric about O, then the area $A(K)$ of K is not greater than 4 [2]. A number of variants of Minkowski's theorem are known, obtained by replacing this symmetry condition by some other condition; see for example [3], [4].

Suppose now that instead of requiring symmetry about O, we insist that the circumcentre of K lie at O. What is the maximal value of $A(K)$?

Conjecture 1. The set K of maximal area is shown in Fig. 1. The circumradius R is approximately 1.637 and $A(K) \approx 4.04$.

If instead of symmetry about O, we require that the centroid (centre of gravity) of K lie at O, we obtain a result due to Ehrhart [1]: $A(K) \leqslant 4.5$ with equality only when K is an integral unimodular transform of the triangle in Fig. 2.

Let $w(K)$ denote the minimal width of K, that is, the smallest distance between two parallel support lines of K.

Conjecture 2. If the centroid of K lies at O, then $w(K) \leqslant 3\sqrt{2}/2$, with equality when and only when K is the triangle shown in Fig. 2.

Each of these problems has an obvious n-dimensional analogue; however, even the plane versions appear to present some difficulty.

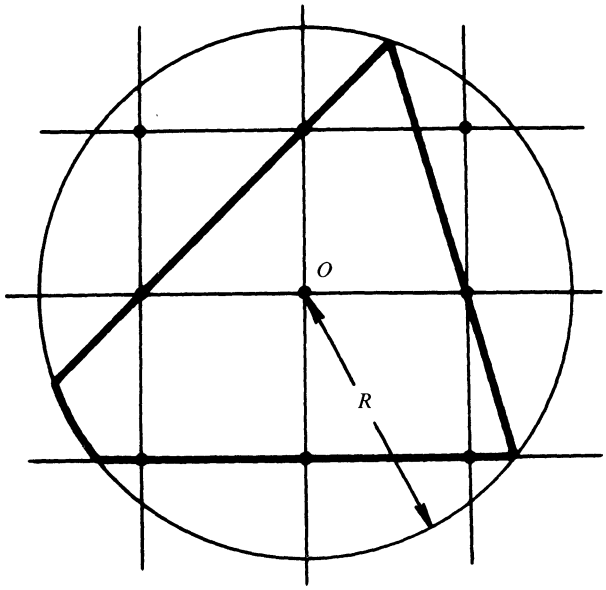


FIG. 1

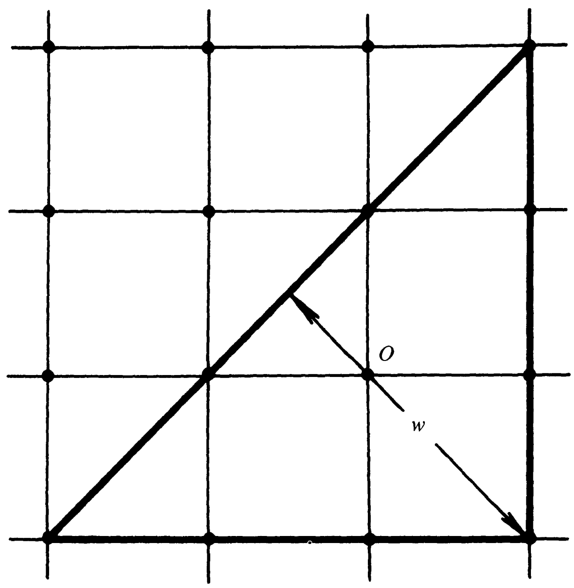


FIG. 2

References

1. E. Ehrhart, Une généralisation du théorème de Minkowski, Comptes Rendus, 240 (1955) 483–485; MR 16-574.
2. H. Minkowski, Geometrie der Zahlen, Leipzig and Berlin (1896); Chelsea, New York, 1953, p. 76.
3. P. R. Scott, On Minkowski's theorem, Math. Mag., 47 (5) (1974) 277; MR 51 # 4050.
4. P. R. Scott, A new extension of Minkowski's theorem, Bull. Austral. Math. Soc., 18 (1978) 403–405; MR 58 # 21952.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

FACTORING REPUNITS

W. M. SNYDER

Department of Mathematics, University of Maine, Orono, Maine 04469

The factorization of a given set of integers into primes has been a problem of interest to mathematicians since antiquity. An example of this is the set of Mersenne numbers, i.e., positive integers of the form $2^n - 1$ where n is a positive integer. In particular, much work has been done in trying to determine which Mersenne numbers are prime and how many such primes there are. Another example is the set of (base ten) "repunits," i.e., integers > 1 in which only ones occur in their decimal expansions. According to Samuel Yates [5], A. H. Beiler coined the term "repunit" since such numbers consist of a string of repeated ones. In this paper, we generalize the notion of repunit as follows: a repunit is any integer > 1 which has a b-adic expansion consisting of only ones for some integer $b > 1$, i.e., a repunit can be represented by $\Sigma_{i=0}^{n} b^i$ for some $b > 1$ and $n > 0$. Notice that all Mersenne numbers > 1 are examples of repunits. There are many interesting problems concerning repunits and much literature exists on the subject. See, for example, [4] and the bibliography in [5]. In particular, the problem of factoring base ten repunits seems to be in the literature since the middle 1800's.

Because of its interest and accessibility, the problem of factoring base ten repunits was brought up in one of the mathematics courses for prospective elementary school teachers at the University of Maine. One of the students, who was interested in finding short cuts to the factoring process, made the following interesting conjecture.

If $R_n(10) = \sum_{i=0}^{n-1} 10^i$ *where* $n > 1$, *then the largest prime divisor of* $R_n(10)$ *is congruent to* 1 *modulo* n. $\quad (\mathrm{I}_{10})$

If we replace the base ten by an arbitrary base $b > 1$, then we obtain the following statement.

If $R_n(b) = \sum_{i=0}^{n-1} b^i$ *where* $n > 1$, *then the largest prime divisor of* $R_n(b)$ *is congruent to* 1 *modulo* n. $\quad (\mathrm{I}_b)$

It seems unlikely that statement (I_{10}) is true for as we shall soon see, there exist n and b for which (I_b) is false.

We now weaken (I_b) as follows.

$R_n(b)$ *has a prime divisor congruent to* 1 *modulo* n. $\quad$ (II)

The purpose of this note is to give an application of algebraic number theory in cyclotomic fields to completely determine for which integers n and b, greater than 1, statement (II) is valid.

THEOREM. *Let n and b be integers greater than* 1. *Let $\Phi_n(X)$ denote the nth cyclotomic polynomial, i.e.,*

$$\Phi_n(X) = \prod_{\substack{a=1 \\ (a,n)=1}}^{n} (X - \zeta^a)$$

where ζ is any primitive nth root of 1. *If there exists a prime q dividing $\Phi_n(b)$ but not dividing n, then $q \equiv 1 \bmod n$.*

The value of this result in connection with (II) lies in the following corollary to the theorem.

COROLLARY. *Under the hypothesis of the theorem, $R_n(b)$ has a prime divisor congruent to* $1 \bmod n$.

Proof of Corollary. First notice that $X^n - 1 = \prod_{d|n}\Phi_d(X)$. Moreover,

$$\frac{X^n-1}{X-1} = \sum_{i=0}^{n-1} X^i = \prod_{\substack{d|n \\ d\neq 1}} \Phi_d(X).$$

Therefore

$$R_n(b) = \prod_{\substack{d|n \\ d\neq 1}} \Phi_d(b)$$

so $\Phi_n(b) \mid R_n(b)$. The corollary now follows immediately from the theorem.

Proof of the Theorem. We factor $\Phi_n(b)$ in the ring R of algebraic integers of $\mathbb{Q}_n = \mathbb{Q}(\zeta)$. Then

$$\Phi_n(b) = \prod_{\substack{a=1 \\ (a,n)=1}}^{n} (b-\zeta^a). \tag{1}$$

We now claim that if A is the ideal in R generated by two distinct factors $b-\zeta^{a_1}$ and $b-\zeta^{a_2}$ given in (1), then either A is R or all the prime ideals dividing A divide the ideal nR. To see this notice that A contains

$$b-\zeta^{a_2}-(b-\zeta^{a_1}) = \zeta^{a_1}-\zeta^{a_2} = \zeta^{a_1}(1-\zeta^{a_2-a_1})$$

and since ζ^{a_1} is a unit in R, we have $1-\zeta^{a_2-a_1} \in A$. There are now two possibilities, which are distinguished by the norm of an element of $\mathbb{Q}_n$. In general if L is a separable extension of degree n of a field K, then there are n distinct K-isomorphisms $\sigma_1 = \mathrm{id}, \sigma_2, \ldots, \sigma_n$ of L into an algebraic closure $\overline{K}$ of K. The norm of $x \in L$ relative to K is $N_{L/K}(x) \equiv \prod_{i=1}^{n}\sigma_i(x)$.

Case 1. Assume that $\zeta^{a_2-a_1}$ is a primitive dth root of 1 where d is a composite. Then $N_{\mathbb{Q}_n/\mathbb{Q}}(1-\zeta^{a_2-a_1}) = 1$ (see, for example [3, p. 73]), in which case $1 \in A$ so $A = R$.

Case 2. Assume that $\zeta^{a_2-a_1}$ is a primitive pth power root of 1 for some prime p (necessarily dividing n). Then $N_{\mathbb{Q}_n/\mathbb{Q}}(1-\zeta^{a_2-a_1}) = p^e$ for some positive integer e (again see [3]), in which case $p^e \in A$ so $A \mid p^eR$. Thus any prime ideal dividing A must then divide nR. This establishes the claim.

Next we claim that no factor $b-\zeta^{a_0}$ in (1) can be a unit in R. If $b-\zeta^{a_0}$ were a unit, then

$$\pm 1 = N_{\mathbb{Q}_n/\mathbb{Q}}(b-\zeta^{a_0}) = \prod_{\substack{a=1 \\ (a,n)=1}}^{n} (b-\zeta^{a_0a}) = \Phi_n(b).$$

But $\Phi_n(b) \neq \pm 1$ for

$$|\Phi_n(b)| = \prod_{\substack{a=1 \\ (a,n)=1}}^{n} |b-\zeta^a| > 1$$

since $|b-\zeta^a| > 1$ for all a.

Now assume $q \mid \Phi_n(b)$ and $q \nmid n$ for some prime q in $\mathbb{Z}$. Let P be any prime ideal in R such that $P \mid qR$ (so $P \nmid nR$). Without loss of generality assume $P \mid (b-\zeta)R$. For any a with $(a,n)=1$, let

σ_a denote the automorphism of $\mathbb{Q}_n$ determined by $\sigma_a(\zeta) = \zeta^a$. Then $\sigma_a(P)$ is again a prime ideal dividing qR (so $\sigma_a(P) \nmid nR$). By the argument above we know that as σ_a ranges over all the $\phi(n)$ automorphisms of $\mathbb{Q}_n$, all the $\sigma_a(P)$ are distinct primes. Therefore q splits completely in R (see, for example [**1**, p. 44]), i.e., $qR = \prod_{i=1}^{\phi(n)} Q_i$ where Q_i are the distinct $\sigma_a(P)$. In general if p is a prime in $\mathbb{Z}$ such that $p \nmid n$, then $pR = \prod_{i=1}^{g} P_i$ where P_i are distinct prime ideals in R and g is determined by $gf = \phi(n)$ where $f = \min\{m \in \mathbb{Z}^+ : p^m \equiv 1 \bmod n\}$ (cf. [**1**] again). Hence in our case $g = \phi(n)$ so $f = 1$ and thus we conclude that $q \equiv 1 \bmod n$.

This establishes the theorem.

As a simple consequence of the theorem and its corollary we have

PROPOSITION. *For all integers n and b greater than* 1, *$R_n(b)$ has a prime divisor which either divides n or is congruent to* 1 mod n.

As a negative result we have the following counterexample to statement (I_b) for $b = 18$. Notice that $R_4(18) = 18^3 + 18^2 + 18 + 1 = 6175 = 5^2 \cdot 13 \cdot 19$ and although 19 is its largest prime factor, $19 \equiv -1 \bmod 4$. The process for obtaining this counterexample is found in the theorem. Namely, we want

$$R_4(b) = \phi_2(b)\phi_4(b) = (b + 1)(b^2 + 1)$$

to be of such a form that $b + 1$ has a large prime factor $\equiv -1 \bmod 4$ whereas $b^2 + 1$ decomposes into smaller (odd) prime factors which by the theorem must be $\equiv 1 \bmod 4$. (We chose $n = 4$ for this is the smallest appropriate n to work with.)

From the theorem and its corollary we see that in order to establish (II) for any given n and b, it suffices to find a prime $q \mid \Phi_n(b)$ but such that $q \nmid n$. However, this is not always possible. For example, consider $R_2(b) = \Phi_2(b) = b + 1$. Then if $b = 2^e - 1$ for some integer $e > 1$, $R_2(b) = \Phi_2(b) = 2^e$. Thus (II) is false for $n = 2$ if and only if $b = 2^e - 1$. On the other hand, we do have the following result.

PROPOSITION. *If p is any odd prime and r any positive integer or if $p = 2$ and $r > 1$, then $\Phi_{p^r}(b)$ is not a power of p.*

Proof. We first consider the special case when $r = 1$. Assume that $\Phi_p(b) = p^a$ for some $a > 1$. (Since $b > 1$, $a = 1$ is not possible.) Since $\Phi_p(b) = \dfrac{b^p - 1}{b - 1}$, we have

$$(b - 1)p^a = b^p - 1. \tag{2}$$

So $b^p \equiv 1 \bmod p$, whence $b \equiv 1 \bmod p$. We write $b = pm + 1$ for some $m \geqslant 1$. Then from (2) we obtain

$$(pm)p^a = (pm + 1)^p - 1$$

so

$$p^a = \sum_{i=1}^{p} \binom{p}{i} (pm)^{i-1}.$$

If p is odd we then have

$$p^a = p + \sum_{i=2}^{p} \binom{p}{i} (pm)^{i-1} = p(1 + ps)$$

for some integer s. But this is impossible.

For the general case, we have that $\Phi_{p^r}(b) = \Phi_p(b^{p^{r-1}})$ (cf. [**2**, p. 206]). Thus the proposition follows from the special case above if p is odd. For $p = 2$, we must insure that $b^{2^{r-1}} \neq 2^e - 1$ for any integer $e > 1$ (by the remark above the proposition). But if $b^{2^{r-1}} = 2^e - 1$, then b would have to be odd and thus since $r > 1$, $b^{2^{r-1}}$ is a square and so $b^{2^{r-1}} \equiv 1 \bmod 4$. But $2^e - 1 \equiv -1 \bmod 4$.

This establishes the proposition.

For technical reasons we prove the following result.

PROPOSITION. *If p is an odd prime and r is any positive integer, then $\Phi_{p^r}(c)$ has a prime factor distinct from p for any integer $c \neq \pm 1, 0$ except for $p = 3$, $r = 1$, and $c = -2$.*

Proof. First assume $r = 1$. Because of the previous proposition, we need only consider $c < -1$. If we assume $\Phi_p(c) = p^a$ where $a > 1$, then the proof in the previous proposition works here. So now assume $a = 1$, i.e., $\Phi_p(c) = p$. Then

$$\Phi_p(c) = \frac{c^p - 1}{c - 1} = \frac{1 + |c|^p}{1 + |c|} > \frac{1 + |c|^p}{2|c|} > \frac{|c|^{p-1}}{2}.$$

Thus if $\Phi_p(c) = p$, then on one hand we must have $|c|^{p-1} < 2p$ from which it easily follows that $|c| \leq p - 1$. On the other hand, since $c \equiv 1 \bmod p$, we have $|c| = p - 1$. But it is easy to see that $(p-1)^{p-1} < 2p$ holds only for $p = 3$. In this case we have $\Phi_3(-2) = (-2)^2 + (-2) + 1 = 3$ which is the exceptional case.

If $r > 1$, then since

$$\Phi_{p^r}(c) = \Phi_p(c^{p^{r-1}}),$$

the result follows by the special case.

This establishes this proposition.

In light of the second proposition and the corollary to the theorem, we now need to investigate the validity of (II) for those n which contain at least two distinct primes. Let us start with an example. Notice that $\Phi_6(2) = 2^2 - 2 + 1 = 3$. However $R_6(2)$ still has a prime divisor congruent to $1 \bmod 6$. For

$$R_6(2) = \Phi_2(2)\Phi_3(2)\Phi_6(2)$$

and

$$\Phi_3(2) = 2^2 + 2 + 1 = 7.$$

This is not a coincidence as the following proposition indicates.

PROPOSITION. *Let $n = 2m$ where m is an odd integer greater than 1. If there exists a prime $q \mid \Phi_m(b)$ such that $q \equiv 1 \bmod m$, then $q \equiv 1 \bmod 2m$.*

Proof. This is an immediate consequence of the Chinese Remainder Theorem for if q is a prime $\equiv 1 \bmod m$, then since $m > 1$, q must be odd, i.e., $q \equiv 1 \bmod 2$.

We are now in a position to characterize all $n, b > 1$ for which (II) is valid.

THEOREM. *$R_n(b)$ has a prime divisor congruent to $1 \bmod n$ if and only if $n \neq 2$ or $n = 2$ and $b \neq 2^e - 1$ for all integers e greater than 1.*

Proof. We need only consider n having at least two distinct prime factors and n either odd or a multiple of 4. These cases will now be handled by the following lemma.

LEMMA. *If n has at least two distinct prime divisors and either $n \equiv 0 \bmod 4$ or n is odd, then $\Phi_n(c)$ has a prime factor not dividing n for any $c \neq \pm 1, 0$.*

Proof of the Lemma. We first establish the lemma for n odd and square-free. In this case we claim that for any prime $p \mid n$, $\Phi_n(c) \mid \Phi_p(c^{np^{-1}})$. This is shown by induction on the number of prime factors of n. Suppose $p \mid n$. Then since n has at least two distinct prime factors, there exists a prime $q \mid n$ such that $q \neq p$. Assume that the claim has been established for n/q, i.e.,

$$\Phi_{n/q}(c) \mid \Phi_p(c^{nq^{-1}p^{-1}})$$

for all integers c. But $\Phi_{n/q}(c^q) = \Phi_n(c)\Phi_{n/q}(c)$ (cf. [2]), whence $\Phi_n(c) \mid \Phi_{n/q}(c^q)$ so

$$\Phi_n(c) \mid \Phi_p(c^{np^{-1}})$$

as desired.

Now suppose $\Phi_n(c) \equiv 0 \bmod p$ for some prime $p \mid n$ and some integer c. Let q be a prime dividing n but $q \neq p$. Then by the above claim we must have $\Phi_q(d) \equiv 0 \bmod p$, where $d = c^{nq^{-1}}$. However, this is impossible for otherwise we would have $0 \equiv \Phi_q(d)(d-1) = d^q - 1 \bmod p$ so that either $d \not\equiv 1 \bmod p$ in which case $q \equiv 0 \bmod p-1$ which is impossible, or $d \equiv 1 \bmod p$ in which case $\Phi_q(d) \equiv \Phi_q(1) \equiv q \equiv 0 \bmod p$ which is again impossible.

Thus the lemma is now established for n odd and square-free.

Now we remove the restriction that n be square-free. Suppose $n = p_1^{e_1} \cdots p_g^{e_g}$ where the p_i are distinct odd primes dividing n. Then

$$\Phi_n(c) = \Phi_{p_1 \cdots p_g}(c^{p_1^{e_1-1} \cdots p_g^{e_g-1}})$$

(cf. [2]). This case now follows from the square-free case. Notice that in this case we have actually shown that $\Phi_n(c)$ is relatively prime to n for all integers $c \neq \pm 1, 0$.

Finally, suppose $n = 2^a m$ where m is odd and greater than 1 and $a > 1$. Then $\Phi_{2^a m}(c) = \Phi_m(-c^{2^{a-1}})$ (cf. [2]) which is either relatively prime to m when m has at least two distinct prime factors or at least has a prime factor not dividing m when m is a prime or a power of a prime. We now need only show that $\Phi_{2^a m}(c)$ is odd. Since $\Phi_m(-c^{2^{a-1}}) \mid \Phi_p(d)$ for some integer d and any $p \mid m$, if $\Phi_n(c)$ were even, then $\Phi_p(d)$ would also be even. But we claim that $\Phi_p(d)$ is always odd for all integers d. For otherwise if $d \equiv 0 \bmod 2$, then $0 \equiv \Phi_p(d) \equiv \Phi_p(0) \equiv 1 \bmod 2$ which is impossible. On the other hand, if $d \equiv 1 \bmod 2$, then $0 \equiv \Phi_p(d) \equiv \Phi_p(1) \equiv p \equiv 1 \bmod 2$ which is again impossible.

This establishes the lemma which in turn yields the theorem.

References

1. G. J. Janusz, Algebraic Number Fields, Academic Press, New York, 1973.
2. S. Lang, Algebra, Addison-Wesley, Reading, MA, 1965.
3. ______, Algebraic Number Theory, Addison-Wesley, Reading, MA, 1970.
4. D. H. Lehmer, Tests for primality by the converse of Fermat's Theorem, Bull. Amer. Math. Soc., vol. 33 (1927) 327–340.
5. S. Yates, Prime Divisors of Repunits, J. Recreational Math., vol. 8(1) (1975) 33–38.

MOST MONOTONE FUNCTIONS ARE NOT SINGULAR

F. S. CATER
Department of Mathematics, Portland State University, Portland, OR 97207

By a singular function we mean a continuous function of bounded variation whose derivative is almost everywhere 0. In [2] Tudor Zamfirescu proved that most continuous monotone functions *are* singular in the sense that the set of singular functions is residual (the complement of a set of first category) in the metric space of all continuous monotone functions under the sup metric. He could not extend this to continuous functions of bounded variation because, under the sup metric, this space is first category in itself. So he extended his result to functions of uniformly bounded variation.

In this note, we study a very natural complete metric on the larger set of all functions of bounded variation. In this space we find that most monotone functions are *not* singular because the set of singular functions is only first category. Indeed it is closed and nowhere dense. More generally, the topology of the space of all continuous functions of bounded variation is essentially the cartesian product topology of the subspace of singular functions with another subspace.

Let B denote the set of real-valued functions of bounded variation on $[0, 1]$, let BC denote the set of continuous functions in B, let M denote the set of nondecreasing functions in B, and let MC denote the set of continuous functions in M. For any partition P: $0 = a_0 < a_1 < \cdots < a_n = 1$, and function f, let $\Sigma_P f$ denote the usual sum $\Sigma_{i=1}^{n} |f(a_i) - f(a_{i-1})|$. The partition Q is a refinement of P if each point in P is a point in Q.

We begin with a lemma that provides somewhat more than we need.

LEMMA 1. *Let f be absolutely continuous on $[0, 1]$ and let g be a real-valued function on $[0, 1]$ such that 0 is a derived number of g at almost every point in $[0, 1]$. Let P be any partition and let $\varepsilon > 0$. Then there is a refinement Q of P such that*

$$\Sigma_Q(f+g) + \varepsilon \geqslant \Sigma_Q f + \Sigma_Q g.$$

Proof. Choose $\delta > 0$ such that if $[a_j, b_j]$ are pairwise nonoverlapping intervals with $\Sigma_j(b_j - a_j) < \delta$, then $\Sigma_j |f(b_j) - f(a_j)| < \frac{1}{4}\varepsilon$. Almost all points x in $[0, 1]$ are endpoints of intervals of the form $[c, d]$ where $|g(d) - g(c)| < \frac{1}{4}\varepsilon(d - c)$ and (c, d) contains no point of P. Indeed these intervals form a Vitali covering of almost all points in $[0, 1]$. We use the Vitali covering theorem to extract countably many of these intervals $[c_i, d_i]$ that are nonoverlapping and $\Sigma_i(d_i - c_i) > 1 - \delta$. Then $\Sigma_i |g(d_i) - g(c_i)| < \frac{1}{4}\varepsilon$. Let Q be the partition consisting of all the c_i and d_i together with all the points of P. Then Q refines P. Let $[r_k, s_k]$ be the intervals determined by Q that are not of the form $[c_i, d_i]$. Then $\Sigma_k(s_k - r_k) < \delta$, so $\Sigma_k |f(r_k) - f(s_k)| < \frac{1}{4}\varepsilon$.

Now

$$\begin{aligned}
\Sigma_Q(f+g) &= \Sigma_k |(f+g)(s_k) - (f+g)(r_k)| + \Sigma_i |(f+g)(d_i) - (f+g)(c_i)| \\
&\geqslant \Sigma_k |g(s_k) - g(r_k)| - \Sigma_k |f(s_k) - f(r_k)| \\
&\quad + \Sigma_i |f(d_i) - f(c_i)| - \Sigma_i |g(d_i) - g(c_i)| \\
&\geqslant \Sigma_Q f + \Sigma_Q g - 2\Sigma_k |f(s_k) - f(r_k)| - 2\Sigma_i |g(d_i) - g(c_i)| \\
&\geqslant \Sigma_Q f + \Sigma_Q g - \varepsilon.
\end{aligned}$$

LEMMA 2. *Let f be absolutely continuous and let g be a function in B with 0 derivative almost everywhere. Then $V(f+g) = V(f) + V(g)$ where V denotes total variation in $[0, 1]$.*

Proof. Choose $\varepsilon > 0$. Let P_1 and P_2 be partitions such that for any refinement Q of $P_1 \cup P_2$, $\Sigma_Q f > V(f) - \varepsilon$, $\Sigma_Q g > V(g) - \varepsilon$. By Lemma 1, there is a refinement Q of $P_1 \cup P_2$ such that

$$V(f+g) \geqslant \Sigma_Q(f+g) \geqslant \Sigma_Q f + \Sigma_Q g - \varepsilon \geqslant V(f) + V(g) - 3\varepsilon.$$

Since ε is arbitrary, we have $V(f+g) \geqslant V(f) + V(g)$. The reverse inequality is clear.

For f and g in B, let $d(f, g) = V(f - g) + |f(0) - g(0)|$. Routine arguments show that d is a complete metric on B. Also $\sup|f - g| \leqslant d(f, g)$.

THEOREM 1. *Let S be the set of functions in B with 0 derivative* a.e., *and let T denote the set of absolutely continuous functions in B vanishing at 0. Then S and T are closed nowhere dense subsets of B and B is homeomorphic to the cartesian product $T \times S$.*

Proof. For h in B put $f(x) = \int_0^x h'(u)\,du$ and $g(x) = h(x) - f(x)$. Then f is in T and g is in S. Identify h with the pair (f, g). In particular, any f in T is identified with $(f, 0)$ and any g in S is identified with $(0, g)$. If also h_1 is identified with the pair (f_1, g_1), then $d(h, h_1) = d(f, f_1) + d(g, g_1)$ by Lemma 2. The rest is clear.

If we had required that all functions in S vanish at 0 and had dropped this requirement for functions in T, the conclusion would have been the same. Thus the set of absolutely continuous functions is also closed and nowhere dense in B. Analogous arguments prove Theorem 1 when B is replaced by M or by BC or by MC. We state

THEOREM 2. *Let S denote the set of singular functions in BC (or MC) and let T denote the set of absolutely continuous functions in BC (or MC) vanishing at 0. Then S and T are closed nowhere dense subsets of BC (or MC), and BC (or MC) is homeomorphic to the cartesian product $T \times S$.*

The set of singular functions in MC has cardinality c and the set of functions in MC that are not singular also has cardinality c. Theorem 2 together with the work in [2] gives rise to a curious result.

THEOREM 3. *There exist complete metrics d and D on the set of real numbers, R, with $d \leqslant D$, such that the set of positive numbers is closed and nowhere dense in (R, D) but is residual in (R, d). Moreover, (R, d) and (R, D) are both arcwise connected.*

LEMMA 3. *Let f, g, h lie in MC and $h = f + g$. Then for any set $E \subset [0,1]$,*

$$\lambda h(E) = \lambda f(E) + \lambda g(E),$$

where λ denotes Lebesgue outer measure.

Proof. Choose $\varepsilon > 0$. Let U, V, W be open sets such that

$$f(E) \subset U, g(E) \subset V, h(E) \subset W, |\lambda U - \lambda f(E)| < \varepsilon, |\lambda V - \lambda g(E)| < \varepsilon, |\lambda W - \lambda h(E)| < \varepsilon.$$

Each set $f^{-1}(U), g^{-1}(V), h^{-1}(W)$ is the union of nonoverlapping intervals and so is their intersection. Let the component intervals of $f^{-1}(U) \cap g^{-1}(V) \cap h^{-1}(W)$ be $(a_i, b_i)_{i=1}^{\infty}$. Now

$$h(b_i) - h(a_i) = f(b_i) - f(a_i) + g(b_i) - g(a_i)$$

for each i, so

$$\Sigma_i(h(b_i) - h(a_i)) = \Sigma_i(f(b_i) - f(a_i)) + \Sigma_i(g(b_i) - g(a_i)).$$

The conclusion follows from

$$\lambda h(E) \leqslant \Sigma_i(h(b_i) - h(a_i)) \leqslant \lambda W,$$
$$\lambda f(E) \leqslant \Sigma_i(f(b_i) - f(a_i)) \leqslant \lambda U,$$
$$\lambda g(E) \leqslant \Sigma_i(g(b_i) - g(a_i)) \leqslant \lambda V.$$

Now we find the distances from any function h in MC to S and T in Theorem 1.

THEOREM 4. *Let h lie in MC and let S and T be as in Theorem 1. Let $s = \sup\{\lambda h(E): E \subset [0,1]$ and $\lambda E = 0\}$. Then $\int_0^1 h'(u)\,du$ is the distance from h to S and $|h(0)| + s$ is the distance from h to T.*

Proof. Say h corresponds to (f, g) in the proof of Theorem 1. Then $d(f, 0) = V(f)$ is the distance from h to S and $d(0, g) = V(g) + |h(0)|$ is the distance from h to T. Also f and g are in MC, so

$$V(f) = \int_0^1 f' = \int_0^1 h' \quad \text{and} \quad V(g) = g(1) - g(0).$$

Let

$$E = \{x: g \text{ does not have a 0 derivative at } x\}.$$

Then $\lambda E = 0$ and $\lambda g([0,1] \setminus E) = 0$ by [1]. Hence $\lambda f(E) = 0$ by [1], and $\lambda g(E) = g(1) - g(0)$. By Lemma 3, $\lambda h(E) = g(1) - g(0) = V(g)$. If $E_1 \supset E$ and $\lambda E_1 = 0$, then we apply [1] and Lemma 3 again to obtain $\lambda h(E_1) = g(1) - g(0) = V(g)$. It follows that $V(g) = s$, and the proof is complete.

We conclude by sketching some results we do not have room to prove here. Let h lie in BC and put $v(x) = V_0^x(h) + |h(0)|$. Then v is in BC and v and h are the same distances from S and T. For any partition P: $0 = a_0 < a_1 < \cdots < a_n = 1$ and set E put

$$s(P, E) = \sum_{i=1}^{n} \lambda h((a_{i-1}, a_i) \cap E).$$

Put $s = \sup\{s(P, E): P$ is a partition and $\lambda E = 0\}$. Then the distance from h to S is $\int_0^1 |h'(u)|\,du$ and the distance from h to T is $|h(0)| + s$. If A is any set, $\lambda v(A) = \sup\{s(P, A): P$ is a partition$\}$.

Another approach to this matter is to identify each function in B with a finite signed Borel measure on $[0, 1]$. Measures in T are absolutely continuous with respect to λ and measures in S are singular with respect to λ. Lemma 2 can be derived from the Lebesgue decomposition of a signed measure.

References

1. E. Hewitt and K. Stromberg, Real and Abstract Analysis, Springer-Verlag, New York, 1965, pp. 269, 288–290.
2. Tudor Zamfirescu, Most monotone functions are singular, this MONTHLY, 88 (1981) 47–49.

ON THE CURL OF A VECTOR FIELD

J.-F. DUMAIS

Département des Sciences Pures, Université du Québec à Rimouski, Rimouski, Québec, Canada G5L 3A1

Modern formulations of advanced calculus have an increased geometrical flavour, derived from an extensive use of the language and concepts of linear algebra. In the present note, we suggest an approach to the concept of curl of a vector field in the same spirit, emphasizing geometrical meaning.

There are several ways of introducing this vector, the most common of which consists in specifying its coordinates with respect to some cartesian frame. But this approach has two drawbacks. The first one is that it completely conceals the geometrical meaning of the curl. The second one is that the invariance of the vector so defined, under changes of coordinates, is not manifest and must be discussed separately; and even after such a discussion, this invariance may remain somewhat of a mystery. This second drawback was pointed out by Wyler [**1**], who gave a clever remedy.

One may also define the curl in terms of a two-form, using exterior calculus [**2**]. This involves identifying the vector field with a one-form, and introducing the curl as the one-form dual (via the $*$-isomorphism) to the exterior derivative of the one-form associated to the field. This approach is undoubtedly elegant and does not suffer from the second drawback mentioned above. However, in introductory courses, one may not wish to, or not have time to introduce the necessary machinery. More importantly, this formulation may not foster an intuitive grasp of the close connection between the curl of a vector field and the local geometry of the vector field.

Finally, one may let the geometrical meaning, or even the definition of the curl emerge from Stokes' theorem. But this may not tell as *directly* and as simply as possible what the curl is all about.

In the present note, we suggest a modernized version of a beautiful old way of introducing the curl, used in many textbooks on vector analysis (see for instance [**3**] and [**4**]), and built around the concept of circulation of the vector field about small circles. This approach has the important advantage of revealing the geometrical meaning of the curl at the outset. Moreover, invariance under changes of coordinates is manifest, and the relation of the curl to the orientation of space is understood geometrically. Finally, the approach fits naturally into the modern linear-algebraic formulation of advanced calculus.

For the present purposes, a vector field will be a differentiable map $\vec{A}: \mathcal{O} \to E_3$, where $\mathcal{O} \subset E_3$ is open. The space E_3 is oriented. Given $\vec{u}, \vec{v} \in E_3$ and $\phi \in E_3^*$, we denote the linear form $\vec{w} \to \vec{u} \cdot \vec{w}$, by $\vec{u}^*$, the linear operators $\vec{w} \to (\vec{u} \cdot \vec{w})\vec{v}$ and $\vec{w} \to \vec{u} \times \vec{w}$, by $\vec{v}\vec{u}^*$ and $\vec{u}^\times$, and the vector associated to the linear form ϕ, by ϕ^* ($\phi(\vec{w}) = \phi^* \cdot \vec{w}$). We recall that any linear operator $L \in L(E_3, E_3)$ decomposes uniquely into a sum $L_+ + L_-$, of symmetric and antisymmetric operators ($L_+^* = L_+$, $L_-^* = -L_-$, "$*$" meaning "adjoint"). Finally we note that the subspace of antisymmetric

operators coincides with the subspace of operators of the form $\vec{u}^{\times}$, $\vec{u} \in E_3$. Given $L \in L(E_3, E_3)$, we shall denote by $\tilde{L}$ the vector thus associated with twice the antisymmetric part of L, that is, the vector $\tilde{L}$ defined by $(\tilde{L})^{\times} = L - L^*$. The "mechanics" of the $L \to \tilde{L}$ map is revealed by the following basic and easily established identity:

$$\widetilde{\vec{v}\,\vec{u}^*} = \vec{u} \times \vec{v}. \tag{1}$$

A vector field may be pictured as a distribution of vectors over its domain $\mathcal{O}$. The simplest example is that of a uniform distribution, for which the same vector is attached to every point. Locally, say in the neighborhood of some point $\vec{a} \in \mathcal{O}$, $\vec{A}$ may be pictured as the superposition of such a constant vector field, equal to $\vec{A}(\vec{a})$, and of a local deviation $\vec{A}'(\vec{a} + \vec{h}) = \vec{A}(\vec{a} + \vec{h}) - \vec{A}(\vec{a})$, which vanishes at $\vec{a}$. It is this deviation $\vec{A}'$ that gives "structure" to the local distribution. For example, $\vec{A}'$ may appear to diverge away from point $\vec{a}$, or to converge towards it, a natural measure of this tendency being the surface integral of the normal component of $\vec{A}'$ over small spheres centered at $\vec{a}$. This leads to the concept of divergence [**5**].

On the other hand, given an axis going through point $\vec{a}$, $\vec{A}'$ may appear to "turn around" that axis. A natural measure of this tendency is the so-called circulation of $\vec{A}'$ about small circles centered at $\vec{a}$ and perpendicular to that axis. If $\Gamma(R, \vec{n})$ is such a circle, of radius R, perpendicular to unit vector $\vec{n}$ and with orientation specified by $\vec{n}$ (i.e., given by any parametrization $\xi \to \vec{r}(\xi)$ such that $\left(\vec{r} \times \dfrac{d\vec{r}}{d\xi}\right) \cdot \vec{n} > 0$), the circulation of $\vec{A}'$ about $\Gamma(R, \vec{n})$ is defined as the line integeral $\int_{\Gamma(R,\vec{n})} \vec{A}'(\vec{r}) \cdot d\vec{r}$. We note that

$$\int_{\Gamma(R,\vec{n})} \vec{A}(\vec{r}) \cdot d\vec{r} = \int_{\Gamma(R,\vec{n})} \vec{A}'(\vec{r}) \cdot d\vec{r}, \tag{2}$$

which reflects the fact that the "uniform part of $\vec{A}$ around $\vec{a}$," $\vec{A}(\vec{a})$, has no bearing on the "rotation tendency." In order to exhibit the behavior of integral (2) at small radius R, it is natural to use the differential of $\vec{A}$ at $\vec{a}$, $d\vec{A}_{\vec{a}}$. Denoting by $\Gamma_0(R, \vec{n})$ the circle obtained by translating $\Gamma(R, \vec{n})$ to the origin, we have

$$\int_{\Gamma(R,\vec{n})} \vec{A}'(\vec{r}) \cdot d\vec{r} = \int_{\Gamma_0(R,\vec{n})} \vec{A}'(\vec{a} + \vec{h}) \cdot d\vec{h} = \int_{\Gamma_0(R,\vec{n})} d\vec{A}_{\vec{a}}(\vec{h}) \cdot d\vec{h} + \int_{\Gamma_0(R,\vec{n})} o(\vec{h}) \cdot d\vec{h}. \tag{3}$$

Owing to the linearity of $d\vec{A}_{\vec{a}}$, it is clear enough that, at fixed $\vec{n}$, the first integral on the right-hand side will be proportional to R^2. What is more remarkable is that it is in fact linear in $R^2\vec{n}$, so that it may be written as $\vec{C} \cdot (\pi R^2 \vec{n})$ for some constant vector $\vec{C}$. Proof of this linearity, as well as motivation for the factor π, will follow shortly. Noting, on the other hand, that

$$\int_{\Gamma_0(R,\vec{n})} o(\vec{h}) \cdot d\vec{h} = o(R^2\vec{n}),$$

we may finally rewrite (3) as

$$\int_{\Gamma(R,\vec{n})} \vec{A}(\vec{r}) \cdot d\vec{r} = \vec{C} \cdot (\pi R^2 \vec{n}) + o(R^2 \vec{n}). \tag{4}$$

Hence, the circulation may be considered a function of $R^2\vec{n}$, and the linear part of this function, which dominates at small R, is determined by the vector $\vec{C}$. This vector is by definition the curl of $\vec{A}$ at $\vec{a}$, and is denoted by $\nabla \times \vec{A}(\vec{a})$. Provided $\vec{C} \neq \vec{0}$, and at R fixed and small enough, the extreme values of the circulation, which differ by sign only, are reached when $\vec{n}$ is at two opposite poles of the unit sphere, close to, or at $\pm|\vec{C}|^{-1}\vec{C}$. On a corresponding meridian, it will vanish at two diametrically opposite positions of $\vec{n}$ close to (or on) the equator.

To see that $\int_{\Gamma_0(R,\vec{n})} d\vec{A}_{\vec{a}}(\vec{h}) \cdot d\vec{h}$ is indeed linear in $R^2\vec{n}$, we may just as easily establish a more general result. Consider a plane Π with unit normal $\vec{n}$ through $\vec{0}$, and within this plane a region Σ

star-shaped around $\vec{0}$, and described by a parametrization $(\lambda, \xi) \to \lambda\vec{r}(\xi)$ where $(\lambda, \xi) \in [0, 1] \times [\xi_1, \xi_2]$ and $\xi \to \vec{r}(\xi)$ is a smooth simple closed curve surrounding the origin and of orientation corresponding to $\vec{n}$. It is easily seen that the area S of Σ is given by

$$S\vec{n} = \tfrac{1}{2}\int_{\xi_1}^{\xi_2}\vec{r}(\xi) \times \frac{d\vec{r}}{d\xi}d\xi. \tag{5}$$

If L is any linear operator on E_3, we then have

$$\begin{aligned}\int_{\partial\Sigma} L(\vec{r}) \cdot d\vec{r} &= \int_{\xi_1}^{\xi_2} L_+(\vec{r}) \cdot \frac{d\vec{r}}{d\xi}d\xi + \int_{\xi_1}^{\xi_2} L_-(\vec{r}) \cdot \frac{d\vec{r}}{d\xi}d\xi \\ &= \int_{\xi_1}^{\xi_2} \frac{d}{d\xi}\left(\tfrac{1}{2}\vec{r} \cdot L_+(\vec{r})\right) d\xi + \int_{\xi_1}^{\xi_2} \tfrac{1}{2}(\tilde{L} \times \vec{r}) \cdot \frac{d\vec{r}}{d\xi}d\xi \\ &= 0 + \tilde{L} \cdot \int_{\xi_1}^{\xi_2} \tfrac{1}{2}\vec{r} \times \frac{d\vec{r}}{d\xi}d\xi\end{aligned}$$

where the first integral vanishes on account of $\vec{r}(\xi_2) = \vec{r}(\xi_1)$. It follows, using (5), that

$$\int_{\partial\Sigma} L(\vec{r}) \cdot d\vec{r} = \tilde{L} \cdot S\vec{n}. \tag{6}$$

This shows that the first integral on the right-hand side of (3) is indeed linear in $R^2\vec{n}$. Moreover, since $\vec{C} = \nabla \times \vec{A}(\vec{a})$ was defined by

$$\int_{\Gamma_0(R,\vec{n})} d\vec{A}_{\vec{a}}(\vec{h}) \cdot d\vec{h} = \vec{C} \cdot (\pi R^2 \vec{n}),$$

we get from (6) the relation between $\nabla \times \vec{A}(\vec{a})$ and $d\vec{A}_{\vec{a}}$:

$$\nabla \times \vec{A}(\vec{a}) = \widetilde{d\vec{A}_{\vec{a}}}. \tag{7}$$

If $\vec{e}_1, \vec{e}_2, \vec{e}_3$ is an orthonormal basis and x_1, x_2, x_3 are associated cartesian coordinates, we have $d\vec{A} = \Sigma_{i=1}^3(\partial\vec{A}/\partial x_i)\vec{e}_i^*$, so that, using (1) and (7),

$$\nabla \times \vec{A} = \sum_{i=1}^{3} \vec{e}_i \times \frac{\partial\vec{A}}{\partial x_i},$$

which gives the usual expression of the curl in terms of coordinates. But, as no coordinates were used originally, the invariance of the curl under coordinate transformations is manifest. However it is *not* invariant under a change of the orientation of the space E_3. For the vector $\vec{u} \times \vec{v}$, when not $\vec{0}$, is partly determined by the requirement that $(\vec{u}, \vec{v}, \vec{u} \times \vec{v})$ be a positively oriented basis. And, upon switching to the alternate choice of positive orientation of E_3, all cross products would be multiplied by -1, so that the orientation of the curve $\Gamma(R, \vec{n})$, for given $\vec{n}$, would be reversed, with the consequence that the left hand side of (4) would change sign, and that the direction of $\vec{C}$ or $\nabla \times \vec{A}(\vec{a})$ would become opposite to what it was. The same conclusion follows from (7), since the map $L \to \tilde{L}$, because it involves the cross product, is orientation dependent. That the direction of the curl should depend on the choice of the orientation of E_3 is not surprising, if we remember that, roughly speaking, the curl tells us around which axis the vector distribution $\vec{A}'$ is "turning" locally, and the sense of that rotation; so this is, in a sense, merely a reflection of the familiar right-hand screw convention.

We may gain further understanding of the geometrical meaning of the curl by elaborating a little on its connection with the local deviation $\vec{A}'$. Denoting by S the symmetric part of $d\vec{A}_{\vec{a}}$, we have

$$\vec{A}'(\vec{a} + \vec{h}) = d\vec{A}_{\vec{a}}(\vec{h}) + o(\vec{h}) = \tfrac{1}{2}\vec{C} \times \vec{h} + s(\vec{h}) + o(\vec{h}). \tag{8}$$

As was stated earlier, it is $\vec{A}'$ that gives structure to the field locally. For small enough $\vec{h}$, when $d\vec{A}_{\vec{a}} \neq \vec{0}$, the distribution $\vec{A}'$ arises essentially from the superposition of $\frac{1}{2}\vec{C}\times\vec{h}$ and $S(\vec{h})$. To visualize the first contribution, we fill up a small neighborhood of point $\vec{a}$ with a family of small coaxial cylinders, with common axis centered at point $\vec{a}$ and parallel to $\vec{C}$. On one of these cylinders, say of radius R, the vectors $\frac{1}{2}\vec{C}\times\vec{h}$ are all tangent to that cylinder and perpendicular to $\vec{C}$, and all have the same length $\frac{1}{2}\,|\,\vec{C}\,|\,R$. They "turn around $\vec{C}$."

Although we are not particularly interested in the second contribution, $S(\vec{h})$, we mention that the associated distribution of vectors may be pictured by noting that, directionwise, they coincide with the normals to the family of quadratic surfaces centered at point $\vec{a}$, of equation $(\vec{r}-\vec{a})\cdot S(\vec{r}-\vec{a}) = \text{constant}$. If S is positive or negative definite, these surfaces are ellipsoids. Otherwise, with ascending rank of S, they are planes, elliptic or hyperbolic cylinders, and hyperboloids of one or two sheets. In either case the resulting distribution shows no tendency to turn around any axis.

Yet another interpretation, for people with a little background in mechanics, may help to understand the curl. Given a unit vector $\vec{n}$, we imagine a little wheel of small radius R, centered at point $\vec{a}$ and with axis constrained to remain parallel to $\vec{n}$. If we pretend that the vector field $\vec{A}$ is a force field, and that it applies forces on the *rim* of the wheel, in such a way that the force on angular section $d\theta$ at $\vec{a}+\vec{\rho}(\theta)$ $(\vec{\rho}\cdot\vec{n}=0)$ is $\vec{A}(\vec{a}+\vec{\rho}(\theta))\,d\theta$, then the moment of this system of forces, about the axis of the wheel, is nothing but the circulation of $\vec{A}$ about the rim of the wheel. Coming back to (4), we see that $\nabla\times\vec{A}(\vec{a})$ tells us how the small wheel is going to spin, in response to this system of forces.

We shall conclude with a look at some standard differential identities involving the curl. Consider two vector fields $\vec{A}$, $\vec{B}$ and a real valued function f, all three differentiable and with common domain $O\subset E_3$. Using standard notations explained below, these identities read

$$\begin{aligned}
\nabla\times(f\vec{A}) &= (\nabla f)\times\vec{A}+f\nabla\times\vec{A}\\
\nabla\cdot(\vec{A}\times\vec{B}) &= -\vec{A}\cdot(\nabla\times\vec{B})+\vec{B}\cdot(\nabla\times\vec{A})\\
\nabla\times(\vec{A}\times\vec{B}) &= (\nabla\cdot\vec{B})\vec{A}-(\vec{A}\cdot\nabla)\vec{B}-(\nabla\cdot\vec{A})\vec{B}+(\vec{B}\cdot\nabla)\vec{A}\\
\nabla(\vec{A}\cdot\vec{B}) &= (\vec{A}\cdot\nabla)\vec{B}-(\nabla\times\vec{B})\times\vec{A}+(\vec{B}\cdot\nabla)\vec{A}-(\nabla\times\vec{A})\times\vec{B}.
\end{aligned}$$

Here $\nabla f(\vec{a})$, the gradient of f at $\vec{a}$, is the vector defined by $df_{\vec{a}}(\vec{h})=\vec{h}\cdot\nabla f(\vec{a})$ for all $\vec{h}$, $\nabla\cdot\vec{A}(\vec{a})=\mathrm{Tr}(d\vec{A}_{\vec{a}})$ is the divergence of $\vec{A}$ at $\vec{a}$, and $(\vec{B}\cdot\nabla)\vec{A}(\vec{a})=d\vec{A}_{\vec{a}}(\vec{B}(\vec{a}))$. To prove these identities, one may introduce coordinates and play around with partial derivatives. But it may be more satisfying to see that these formulas are simple linear-algebraic consequences of the following more basic and obvious identities:

$$\begin{aligned}
d(f\vec{A}) &= \vec{A}(\nabla f)^*+f\,d\vec{A}\\
d(\vec{A}\times\vec{B}) &= \vec{A}^{\times}\,d\vec{B}-\vec{B}^{\times}\,d\vec{A}\\
d(\vec{A}\cdot\vec{B}) &= \vec{A}^*\,d\vec{B}+\vec{B}^*\,d\vec{A}.
\end{aligned}$$

To this end, we use (1) together with the following two identities concerning the trace and antisymmetric part of an operator of the form $\vec{u}^{\times}L$, which are both easy consequences of (1) established below,

$$\mathrm{Tr}(\vec{u}^{\times}L) = -\vec{u}\cdot\tilde{L} \tag{9}$$

$$\widetilde{\vec{u}^{\times}L} = \mathrm{Tr}(L)\vec{u}-L(\vec{u}), \tag{10}$$

and get these short coordinate-free proofs:

$$\begin{aligned}
\nabla\times(f\vec{A}) &= \widetilde{d(f\vec{A})} = \widetilde{\vec{A}(\nabla f)^*}+f\widetilde{d\vec{A}} = \nabla f\times\vec{A}+f\nabla\times\vec{A}\\
\nabla\cdot(\vec{A}\times\vec{B}) &= \mathrm{Tr}\big[d(\vec{A}\times\vec{B})\big] = \mathrm{Tr}(\vec{A}^{\times}\,d\vec{B}-\vec{B}^{\times}\,d\vec{A}) = -\vec{A}\cdot(\nabla\times\vec{B})+\vec{B}\cdot(\nabla\times\vec{A})
\end{aligned}$$

$$\begin{aligned}
\nabla \times (\vec{A} \times \vec{B}) &= \widetilde{d(\vec{A} \times \vec{B})} = \widetilde{\vec{A}^{\times}\, d\vec{B}} - \widetilde{\vec{B}^{\times}\, d\vec{A}} \\
&= \operatorname{Tr}(d\vec{B})\vec{A} - d\vec{B}(\vec{A}) - \operatorname{Tr}(d\vec{A})\vec{B} + d\vec{A}(\vec{B}) \\
&= (\nabla \cdot \vec{B})\vec{A} - (\vec{A} \cdot \nabla)\vec{B} - (\nabla \cdot \vec{A})\vec{B} + (\vec{B} \cdot \nabla)\vec{A} \\
\nabla(\vec{A} \cdot \vec{B}) &= [d(\vec{A} \cdot \vec{B})]^* = [\vec{A}^*\, d\vec{B} + \vec{B}^*\, d\vec{A}]^* = (d\vec{B})^*(\vec{A}) + (d\vec{A})^*(\vec{B}) \\
&= \left[d\vec{B} - \widetilde{(d\vec{B})^{\times}}\right](\vec{A}) + \left[d\vec{A} - \widetilde{(d\vec{A})^{\times}}\right](\vec{B}) \\
&= (\vec{A} \cdot \nabla)\vec{B} - (\nabla \times \vec{B}) \times \vec{A} + (\vec{B} \cdot \nabla)\vec{A} - (\nabla \times \vec{A}) \times \vec{B}.
\end{aligned}$$

To prove (1) we note that, for any $\vec{w}$,

$$\widetilde{\vec{v}\vec{u}^*} \times \vec{w} = (\vec{v}\vec{u}^* - \vec{u}\vec{v}^*)(\vec{w}) = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{v} \cdot \vec{w})\vec{u} = (\vec{u} \times \vec{v}) \times \vec{w}.$$

By linearity, it is enough to check (9) and (10) for operators of the form $\vec{w}\vec{v}^*$. We find, using (1),

$$\operatorname{Tr}[\vec{u}^{\times}(\vec{w}\vec{v}^*)] = \operatorname{Tr}[(\vec{u} \times \vec{w})\vec{v}^*] = \vec{v} \cdot (\vec{u} \times \vec{w}) = -\vec{u} \cdot (\widetilde{\vec{w}\vec{v}^*}),$$

and

$$\begin{aligned}
\widetilde{\vec{u}^{\times}(\vec{w}\vec{v}^*)} &= \widetilde{(\vec{u} \times \vec{w})\vec{v}^*} = \vec{v} \times (\vec{u} \times \vec{w}) \\
&= (\vec{v} \cdot \vec{w})\vec{u} - (\vec{v} \cdot \vec{u})\vec{w} = \operatorname{Tr}(\vec{w}\vec{v}^*)\vec{u} - (\vec{w}\vec{v}^*)(\vec{u}).
\end{aligned}$$

An interesting by-product of Eq. (10) is the following formula for the image of the cross product of two vectors by a linear operator, which the author has not met elsewhere:

$$L(\vec{u} \times \vec{v}) = \operatorname{Tr}(L)\vec{u} \times \vec{v} - L^*(\vec{u}) \times \vec{v} - \vec{u} \times L^*(\vec{v}).$$

This formula is obtained by taking the cross product of each side of (10) with $\vec{v}$ and replacing L by L^*.

References

1. O. Wyler, Invariant definitions for vector calculus, this MONTHLY, 75 (1968) 394–396.
2. L. H. Loomis and S. Sternberg, Advanced Calculus, Addison-Wesley, Reading, MA, 1968.
3. D. E. Rutherford, Vector Methods, Oliver and Boyd, Edinburgh and London, 1965, p. 63.
4. M. Schwartz, S. Green, and W. A. Rutledge, Vector Analysis, Harper, New York, 1960, pp. 127–131.
5. S. Lang, Calculus of Several Variables, Addison-Wesley, Reading, MA, 1973, p. 331.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

A MODIFIED MOORE METHOD FOR TEACHING UNDERGRADUATE MATHEMATICS

DAVID W. COHEN

Department of Mathematics, Smith College, Northampton, MA 01063

1. Introduction. The late R. L. Moore made famous a method of teaching mathematics which has had a profound influence on educators [3]. Professor Moore's most important departure from age-old teaching practices was that he gave no lectures. Instead he gave his students theorems to prove and required them to work on the proofs until they were correct. Professor Moore selected

those who would take his graduate course from a list of students eager to participate, and many of his students became excellent productive mathematicians.

When the "Moore method" was attempted by others at the undergraduate level, however, the results often were disappointing. Instructors of undergraduates usually cannot handpick their students. Moreover, undergraduates normally have not had the experience proving theorems and writing mathematics necessary for meaningful progress without help from an instructor. Finally, the Moore method sometimes creates an unhealthy atmosphere of competition and isolation among students.

Recently, Professor Moore's philosophy of shifting the focus from instructor to student has been incorporated in a wide variety of teaching methods, which are more successful for undergraduate courses. (See [4] and [5].) The method I describe here has worked very well and is the most comprehensive program I have seen for introducing a Moore-type method to undergraduate teaching.

Although this Modified Moore method was developed to teach mathematics, it is based on principles that apply to the teaching of most subjects.

1. Students understand better and remember longer what they discover themselves than what is told to them.
2. People master an idea most thoroughly when they teach it to someone else.
3. Effective writing and clear thinking are inextricably linked.

The following objections usually arise when one thinks about teaching methods devised to put these principles into practice.

1. The truths taught in most courses took great thinkers many years to discover. We cannot expect our students to discover them all in one semester. Further, the amount of the material we cover in our courses is at a minimum and should not be reduced.
2. It might be wonderful training for students to try to teach what they have just learned, but pity the poor student who must learn from such a teacher.
3. Yes, we wince at some of the writing we get from students, but we are not teaching writing courses. We do not have the time or the training to teach writing. The serious students will learn to write when they get to graduate school and read journals. If their writing is fuzzy, we can test their understanding with short-answer questions.

The Modified Moore method, based on the principles stated above, answers these objections. In brief, the method divides a class into small groups, each of which is responsible for a weekly question. Over the course of a week each small group will study and answer the question, write a short paper presenting the answer, and prepare to teach the question and its answer to the rest of the class. The questions and their answers contain all the ideas covered in the course.

2. Preparation and Scheduling for the Course. The first task for the instructor is to break the course material into a list of *questions*. These may resemble questions asked on a take-home examination at the end of the course. Some examples of questions for a course in real analysis are provided in the Appendix. An average student with some coaching from the instructor should be able to understand a question and write an answer to it in one week. There should be enough questions so that any student who can answer all of them thoroughly will have mastered all the material for the course. The number of questions should be about two or $2\frac{1}{2}$ times the number of weeks alloted for the course. This number will provide weekly allotments of two or three questions per week and allow three weeks to be reserved for organizing the class and giving examinations.

The next task is to compile a list of *basics* that the students must understand to answer the questions. These are usually definitions or axioms or theorems from other courses. In the

[Continued on p. 487.]

Appendix I have provided examples of basics as well as questions.

With the Modified Moore method the class is scheduled to meet as an entire class once a week for *class meetings* of 90 to 110 minutes. It does not work well to substitute two 50-minute class meetings for the one longer meeting, since ample time is required for students to ask questions about small points. Furthermore, the questions for the week are often intimately related and are better discussed in a single meeting. Two other hours on two separate days each week are then scheduled for *coaching sessions.*

3. Organizing the Class and Assigning Work. The first day of class is devoted to a thorough explanation of the method, its objectives, and what is expected of the students.

The class is then divided into teams of two or three students each. Teams of three work best. These do not change until midterm when all teams must change for the second half of the term. New teams are assigned to avoid those consisting only of weak or only of strong students.

At the end of each class meeting the instructor assigns questions for the next class meeting. Several teams may be assigned the same question. After assigning the questions, the instructor announces the schedule for the first coaching session.

The assignment is to work as a team to: 1) understand the question, 2) find the answer, and 3) write a short paper presenting the question and its answer. This paper will be duplicated and distributed to all members of the class. In addition, each team must be prepared to explain to the class its question and the answer it has found. All teams are expected to be prepared for each class meeting, although only one team per question will be called upon to make an oral presentation.

4. The Coaching Sessions. The first coaching session follows the class meeting by one or two days. Each team is required to come during that part of the first coaching session dealing with its question. If several teams are dealing with the same question, all of those teams come at the same time. Each team is expected to have met once prior to this session to try to understand the basics and the question and to formulate ideas about the answer.

When students have discovered the answer before the session, the instructor simply listens to the team's answer and provides whatever help is necessary to make the team's explanation very clear. When a team comes in a state of confusion, on the other hand, the instructor must do some explaining. In the 20 or 30 minutes devoted to each question, he must try to lead the students to the heart of the matter, using a Socratic style as long as time permits it. Sometimes this means asking one or two questions to help students understand how they almost had the idea themselves. At other times it means giving a complete explanation, and hoping that the students will make a discovery on some other questions. I try to make sure that at least one member of each team has a basic understanding before the teams leave.

The second coaching session follows the first by three days or more. During those days the students prepare drafts of their papers. The teams meet and discuss the work and a team might even agree on a rough written draft of its paper. Each member must then write his or her own draft without consultation.

All teams gather for the second coaching session during which students exchange their drafts with their teammates for peer review. The reviewers pay close attention to content, clarity and style, and make suggestions for improvement.

While this peer review is taking place, the instructor selects drafts at random and reviews them in front of the team. This provides at least one team member with professional comment on a draft, and at the same time gives each member a chance to watch how the instructor reviews. I reserve some time to speak to the class as a whole about common errors or misconceptions found during my reviews.

5. The Weekly Papers. Each paper must consist of a clear, concise, cogent answer to the team's question for the week, written in a style appropriate to the subject for the course, and with careful attention to syntax, punctuation and correct usage of notation commonly used by

professionals in the subject.

Each paper must begin with an abstract of two or three sentences revealing the key idea in the paper. Students perform a task important for clear understanding when they distill a long argument into two or three essential sentences.

Each team decides how the paper is to be written. Sometimes the students divide a paper into sections with each member responsible for writing a section. Sometimes a single student writes the entire paper; however, that student may not write another one until the other members of the team have also written an entire paper.

6. The Class Meetings. At the beginning of the class meeting each team passes a copy of its paper out to all students. Two copies go to the instructor, one to be corrected and returned. The team called upon for the first question for the week then makes its presentation. Each team must plan its presentation and allow time for questions from the class. Usually each member is responsible for one part of the presentation.

The instructor's role during the presentation is a delicate matter. If there is a point that seems unclear to me, I wait for a student to ask about it; but if no one does, then I ask. I am willing to help a team clarify an explanation, but only after I sense that no one else can do it. With this method students are much less willing to let the instructor dominate discussion than in more conventional classes. They have invested much time and energy in their preparation, and are anxious to explain their ideas clearly to their classmates.

7. Evaluating Student's Work. The instructor should try to read each team's paper carefully and return the papers the day after they are received. He should correct grammar and style as well as content.

The Modified Moore method calls for two comprehensive written examinations; one at midterm and one at the end of the course. Students take these as individuals, not as teams. Everyone is responsible for all the material discussed in class meetings, including material the instructor may have provided on the spur of the moment to clarify or amplify an idea.

8. The Method at Smith College. I have successfully used the Modified Moore method to teach courses at all levels at Smith: a course on "The Infinite" for nonmathematics majors, a course on Hilbert spaces for senior mathematics honors students, as well as standard intermediate-level courses. The method has worked well with classes having as few as 5 students and as many as 25. I suspect that it would not work very well for classes much over 25 because the coaching sessions would be too large to allow time for the Socratic approach.

Students are encouraged to consult appropriate text books, which I place on reserve at the library. Sometimes I even cite the pages of a book that contain the answer to a question for the week. To read the answer a student must understand the author's notation and refer back to results the author has proved earlier to see if those results have been proved in our class.

I grade papers and oral presentations either "0", "1", or "2". Most grades are "1", which indicates satisfactory work. A grade of "2" indicates an unusually impressive piece of work. A grade of "0" indicates that the paper missed the idea altogether. Every member of a team receives the same grade for that team's paper, but members receive individual grades for their part in an oral presentation.

9. Conclusions. The objections cited in section 1 can now be refuted.

The Modified Moore method incorporates the principle of student discovery without undue sacrifice in amount of material covered. While each student grapples each week with only one question, other students are dealing with other questions, so that the total amount of material covered by the class in one week is roughly equal to that which would be covered by lectures. I am convinced by students' performance on examinations that the total amount of material students master in courses taught by the Modified Moore method is at least as great as the amount they master in lecture courses.

The method effectively raises the level of communication between students. Since all students are responsible for all questions on examinations, the presenting teams try hard to teach and the other students demand the clarity they need in order to learn.

Since students write every week and receive prompt evaluation of their writing, they make substantial improvement. They also learn to appreciate the connection between understanding a subject and writing about it.

Furthermore, most students respond well to the responsibility placed on them by the Modified Moore method. They are willing to work hard for a teacher who pays attention to their thinking and writing. I find that the percentage of students who drop courses taught by the Modified Moore method because they are burdensome is no higher than the percentage who drop other courses for that reason. Most students are stimulated by the change from passive to active learning.

Appendix. Below is a sample syllabus and the basics and questions for class meeting number six for the course Introductory Real Analysis taught for four years at Smith College by the Modified Moore method. The course is for sophomore and junior mathematics majors. The prerequisites are Multidimensional Calculus and Linear Algebra. There are ten weeks available for class meetings out of the 13-week semester and the total number of questions for the course is 27.

Week		
1. Orientation		
2. Class meeting	1:	Metrics; completeness of the reals; open and closed subsets of the reals.
3. Class meeting	2:	Limit points; Bolzano-Weierstrass Theorem.
4. ____	3:	Continuity of functions, introduction.
5. ____	4:	Continuity of functions, sums, products, compositions, examples.
6. ____	5:	Differentiability of functions.
7. ____	6:	Cauchy sequences of reals; pointwise and uniform convergence of sequences of functions.
8. Examination		
9. Class meeting	7:	Theorems on uniformly convergent sequences of functions; Weierstrass *M*-Test.
10. ____	8:	Analytic functions; Taylor's Theorem.
11. ____	9:	Lebesgue measure, introduction.
12. ____	10:	Lebesgue integral, introduction.
13. Examination		

Basics for Class Meeting Six. (Week 7)

DEFINITIONS: a Cauchy sequence of real numbers, the limit of a sequence of real numbers, the pointwise limit of a sequence of functions, the uniform limit of a sequence of functions.

Questions for Class Meeting Six. (Week 7)

1. Show that the sequence of functions $f_n(x) = x^n$ converges pointwise but not uniformly on $[0, 1]$ to the function

$$h(x) = \begin{cases} 0 & 0 \leqslant x < 1 \\ 1 & x = 1. \end{cases}$$

2. Show that the sequence of functions $f_n(x) = 1 - x^n$ converges uniformly on $[-p, p]$ for any p with $|p| < 1$.

References

1. J. Gersting and J. Kuczkowski, Why and how to use small groups in the mathematics classroom, Two-year Coll. Math. J., 8 (1977) 270–74.

2. P. R. Halmos, E. E. Moise, and G. Piranian, The problem of learning to teach, this MONTHLY, 82 (1975) 466–476.

3. F. Burton Jones, The Moore method, this MONTHLY, 84 (1977) 273–8.

4. J. A. Murtha, The apprentice system for math methods, this MONTHLY, 84 (1977) 473–6.

5. Warren Page, A small group strategy for enhanced learning, this MONTHLY, 86 (1979) 856–8.

6. J. Weinglass, Small groups: an alternative to the lecture method, Two-year Coll. Math. J., 7 (1976) 15–20.

HOW TO CONSTRUCT AND ANALYZE PROOFS—A SEMINAR COURSE

ROBERT B. REISEL

Department of Mathematical Sciences, Loyola University, Chicago IL 60626

In order to learn mathematics, it is necessary to be able to understand proofs. Mathematics students or graduates who do not know how to construct proofs or how to read them critically have, at best, an incomplete understanding of any mathematics they are studying. A major goal in teaching mathematics is, therefore, to teach proofs and proof techniques. The corresponding problem is when and how to do this.

Traditionally, students first studied proofs in high school geometry courses. They still do, although now proofs are also found in high school algebra courses. The trouble is that rigorous proofs are slighted (and rightly so, I believe) in the usual college calculus courses. By the time students study abstract mathematics, they have forgotten whatever they did know about proofs and are at a loss when asked to write or to analyze a proof. Instructors must take time from the actual mathematical content of their courses to teach proof construction and criticism. With the increase in the amount of mathematics and computer science in the curriculum, however, it is difficult to justify taking time to teach techniques of theorem-proving by omitting important topics.

At Loyola University we decided to solve this problem by running a seminar for sophomore mathematics and computer science majors which would have as its goal the teaching of proof construction and criticism. In the sophomore year these students have a reasonable degree of mathematical maturity from having finished our calculus sequence, but they have not begun to study abstract algebra or real analysis. We decided on a seminar rather than on a more traditional course so that the students will get actual practice in constructing and analyzing proofs. Individual assignments of proofs to be constructed are given to the students who then present their results orally before the seminar. Since it is to be expected that some of the proofs will be incorrect or incomplete, the other students in the seminar must be alert to raise questions and objections. This helps develop a critical sense in the students more effectively than the situation in which students only raise questions about correct proofs given in their texts or by their instructors. An extra bonus of the seminar approach is that it gives the students the experience of learning to think on their feet by trying to answer questions or make corrections on the spot.

We were next faced with the problem of what topic should be studied in the seminar. It should be one that the students recognize as something important and useful. In fact, it should be one that the students find so interesting that they will not only be thinking about the goal of learning to construct proofs, but also will be studying the subject for its own sake. The topic should not require much prior knowledge or experience, the concepts should be clear, and the proofs should be relatively short and straightforward. We finally chose the elementary parts of the theory of metric spaces. Not only does this subject have the above attributes, but also it leads directly from calculus to more rigorous courses in analysis and topology.

Since we were not able to find a book that met our needs, we prepared a set of seminar notes (see [1]). Although other topics could be chosen, I will give a brief description of these notes so that the general level and organization of the seminar can be seen. The first chapter contains a brief explanation of those parts of logic that are needed in proofs. This is not a theoretical

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, T(13-14), S, L. New Tertiary Mathematics. C. Plumpton, P.S.W. MacIlwaine. Pergamon Pr, 1980, $67 set (P). V. 1, Part 1: Pure Mathematics: The Core. xix + 401 pp [ISBN: 0-08-021643-9]; V. 1, Part 2: Basic Applied Mathematics. xiv + 229 pp [ISBN: 0-08-021645-5]; V. 2, Part 1: Further Pure Mathematics. xix + 401 pp [ISBN: 0-08-021644-7]; V. 2, Part 2: Further Applied Mathematics. xii + 221 pp. [ISBN: 0-08-025026-2] "Pure" topics covered are, approximately, elementary calculus, linear algebra, differential equations, and a little abstract algebra, arranged as a single course--thus the order of presentation is not standard in the U.S. "Applied" topics include statics, dynamics, probability, and statistics, the first two in considerable depth. The first volume of each part covers prerequisites for university study in Britain. Many worked-out examples. Valuable for comparison and reference. PZ

General, P. Visiting Scholars' Lectures--1980. John T. White. Math. Ser., No. 14. Texas Tech U, 1981, 139 pp, $10 (P). Six papers on analysis and algebra from Texas Tech's visiting mathematician program. Authors include I. Reiner, I. Juhâsz and J. Wermer, among others. LAS

Elementary, T(13). College Mathematics for Business and the Social Sciences. Arthur Lieberman. Brooks/Cole Pub, 1982, x + 598 pp, $19.95. [ISBN: 0-8185-0474-9] Elementary text serving up a smorgasbord of ideas (sets, linear algebra, math of finance, differential and integral calculus) said to have applications in business. AWR

Elementary, T(13: 1). Mathematics for Elementary School Teachers. Harvey Gerber. Saunders Coll Pub, 1982, xi + 546 pp, $21.95. [ISBN: 0-03-58326-8] Standard presentation of topics for prospective elementary school teachers. Content is supported and motivated by illustrations adapted from current elementary mathematics series. Problem sections are more interesting than the usual for such texts. JJ

Precalculus, T*(13: 1). College Algebra. Michael Sentlowitz, Margaret Trivisone. Addison-Wesley, 1982, xii + 545 pp, $14.95. [ISBN: 0-201-06626-2] Includes the standard topics found in a college algebra textbook (no trigonometry). Features many worked out examples, lots of problems, historical notes and an emphasis on graphing. The writing style makes the book clear and easy to read. CEC

Precalculus, T?(13: 1), S. Basic Mathematics: A Precalculus Course for Science and Engineering. Abdullah H. Al-Moajil, Abdelali Benharbit. Wiley, 1981, xii + 308 pp, $35. [ISBN: 0-471-27941-2] Designed for students in the Middle East, Africa and other Third World countries. Contains a smattering of college algebra, trigonometry, analytic geometry, vectors in R^2, matrices, limits and even derivatives and integrals. The last two topics should be omitted and the others covered more thoroughly. JNC

Precalculus, T(13: 1). College Algebra. Thomas W. Hungerford, Richard Mercer. Saunders Coll Pub, 1982, xiv + 493 pp, $20.95. [ISBN: 0-03-059521-5] Traditional material suitable for use in precalculus and terminal courses for students with two years of high school mathematics. Informal, sound, and student-oriented. Strong features include flexibility, graded exercises, "Do it Yourself" sections on non-essential topics, and an up-front set of guidelines for the student. Book equals authors' Algebra and Trigonometry minus two chapters on trigonometry. JK

Precalculus, T(13: 1). Precalculus Mathematics, Second Edition. Daniel D. Benice. Prentice-Hall, 1982, ix + 501 pp, $21.95. [ISBN: 0-13-694976-2] This revision of the former Precalculus, Algebra and Trigonometry (TR, November 1976) now includes synthetic division; an expanded treatment of the exponential function and natural logarithms; and chapter review exercises as well as new applications, examples and exercises. JNC

Precalculus, T(13: 2), S. Essentials of Technical Mathematics, Second Edition. Richard S. Paul, M. Leonard Shaevel. Prentice-Hall, 1982, xii + 703 pp, $21.95. [ISBN: 0-13-288050-4] Precalculus for

engineering students; new topics include evaluating formulas, SI units, determinants, polar coordinates and linear interpolation. Exercise sets and applications have been expanded and the slide rule appendix has been replaced by one on calculators. (First Edition, TR, March 1975.) JNC

Education, S(17), P*. Changing School Mathematics: A Responsive Process. Ed: Jack Price, J.D. Gawronski. NCTM, 1981, viii + 229 pp, $15. [ISBN: 0-87353-184-1] A series of essays which focuses on the process of change and how to change current mathematics programs to reflect NCTM's Agenda for Action: Recommendations for School Mathematics of the 1980's (TR, December 1980). Excellent addition to NCTM's series of professional references; highly recommended for mathematics coordinators, curriculum developers, and professional mathematics educators. JJ

Education, P? CAI Sourcebook: Background and Procedures for Computer Assisted Instruction in Education and Industrial Training. Robert L. Burke. Prentice-Hall, 1982, x + 210 pp, $14.95. [ISBN: 0-13-110155-2] The author's goal is to provide a step-by-step approach for the development of microcomputer-based computer-assisted instruction (CAI) courseware, based upon his experiences with large-scale CAI efforts. The book includes a checklist for CAI courseware review, task-analysis forms, samples of frame designs, and an extensive glossary of CAI terms. Unfortunately, it is also pedantic and verbose. JJ

Education, P. Ideas from the Arithmetic Teacher, Grades 1-4: Primary. George Immerzeel, Melvin Thomas. NCTM, 1982, v + 120 pp, $5.40 (P). [ISBN: 0-87353-189-2] A collection of reprints from the IDEAS section of the Arithmetic Teacher. Appropriate for students in grades 1-4, the articles involve numeration, computation, geometry, measurement, and problem-solving. The text complements NCTM's previous collection for grades 4-8. (See TR, December 1979.) JJ

Education, P. Geometric Selections for Middle School Teachers (5-9): The Curriculum Series. Douglas B. Aichele, Melfried Olson. NEA, 1981, 95 pp, $7.95 (P). [ISBN: 0-8106-1720-X] Presentation of topics in geometry which are to be adapted by the reader for use in the middle school mathematics classroom. Topics include axiomatic systems, distance, congruence, constructions, and transformational geometry. Supported by interesting exercises, useful activities, and references. JJ

Graph Theory, T(17: 1), S, P, L. Graph Theory and Applications. H.N.V. Temperley. Math. & its Appl. Halsted Pr, 1981, 130 pp, $44.95. [ISBN: 0-85312-252-0] This book starts with first principles of enumerative graph theory and leads up to various areas of application including game theory and statistical mechanics. The approach is sophisticated. Includes problems at the end of each chapter and a useful list of references. CEC

Graph Theory, P. Algebraic Methods in Graph Theory. Ed: L. Lovász, Vera T. Sós. North-Holland, 1981, $134 set [ISBN: 963-8021-40-3]. V. I, 443 pp; V. II, 400 pp. Proceedings of an August 1978 international conference in Szeged on applications of matroids, groups, eigenvalues, categories, etc., to graph theory. LAS

Combinatorics, S(18), P, L. Mathematics for the Analysis of Algorithms. Daniel H. Greene, Donald E. Knuth. Progress in Comp. Sci., V. 1. Birkhauser Boston, 1981, 107 pp, $10. [ISBN: 3-7643-3046-5] Assembled handouts from Stanford's course Computer Science 255, including homework assignments, mid-term and final exams (with solutions). Covers binomial identities, recurrence relations, operator methods, and asymptotic analysis in a terse yet comprehensive manner. Assumes familiarity with a wide variety of mathematics, notably complex and combinatorial analysis. LAS

Algebra, T(15: 1), S, L*.** A Book of Abstract Algebra. Charles C. Pinter. McGraw-Hill, 1982, xv + 351 pp, $23.50. [ISBN: 0-07-050130-0] The standard introductory topics are covered, i.e., groups, rings and fields culminating with a proof of the unsolvability of the quintic. A particularly clear writing style, which conveys a strong sense of the relevance of the subject, and an outstanding collection of exercises make for an exciting new algebra text. CEC

Finite Mathematics, T(13: 1). Finite Mathematics, Second Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1982, xv + 452 pp, $17.95. [ISBN: 0-673-15536-6] Topics include: linear models, matrices, graphical and simplex methods of linear programming, probability and statistics, and the mathematics of finance. (Portions of this text were previously included in Mathematics with Applications in the Management, Natural and Social Sciences, Second Edition. TR, April 1979.) Study guide and solutions guide are available. JNC

Calculus, T(13-14: 3). Calculus with Analytic Geometry, Second Edition. Nathan O. Niles, George E. Haborak. Prentice-Hall, 1982, xvi + 635 pp, $24.95. [ISBN: 0-13-112011-5] Additions include problems similar to illustrative examples which they immediately follow and warnings to emphasize certain notions or point out common student errors. The calculus of the trigonometric functions is begun by differentiating the tangent function. The treatment remains informal and intuitive with emphasis on how rather than why. (TR, First Edition, April 1971.) JK

Calculus, T*(13: 2, 3), L. Calculus, One and Several Variables with Analytic Geometry, Fourth Edition. S.L. Salas, Einar Hille. Wiley, 1982, xvi + 1136 pp, $33.95. [ISBN: 0-471-04660-4] This edition includes an expanded treatment of three-dimensional calculus and new chapters on complex numbers and differential equations. In addition, simpler proofs, additional examples and exercises, and more detailed explanations have been included in many chapters. (TR, First Edition, June-July 1971; Second Edition, May 1975; Third Edition, June-July 1978 and December 1978.) CEC

Calculus, T(13: 3). Calculus and Analytic Geometry. James E. Shockley. Saunders Coll Pub, 1982, xii + 1199 pp, $32.95. [ISBN: 0-03-018886-5] Another standard 4.5 lb. presentation (ε-δ proofs are incorporated throughout) with chapters on vector analysis and differential equations; primarily physical science applications. JNC

Real Analysis, T(14-16: 1), S. Functions of Several Variables. B.D. Craven. Chapman & Hall, 1981, viii + 136 pp, $9.95 (P); $16.95. [ISBN: 0-412-23340-1; 0-412-23330-4] A brief and brisk text, with exercises and many solutions, on multivariable calculus through Stokes' theorem for differential forms on m-cubes in R^n. Stated prerequisites: elementary calculus and some linear algebra; but this would be a challenging text at the intermediate level. Covers inverse and implicit function theorems and chain rule (rigorously); Lagrange and Kuhn-Tucker conditions; some theory of differential forms; definition of integrals on manifolds (in Appendix). PZ

Real Analysis, T(16-17: 1), S, P, L. Introduction to Approximation Theory, Second Edition. E.W. Cheney. Chelsea Pub, 1982, x + 259 pp, $14.95. [ISBN: 0-8284-0317-1] Virtually unchanged from First Edition (TR, October 1967). A bargain purchase if only for the thirty-one pages of historical notes and references which, unfortunately, end with 1965. JK

Complex Analysis, P. Complex Analysis in Locally Convex Spaces. Seán Dineen. Math. Stud., V. 57. North-Holland, 1981, xiii + 492 pp, $51 (P). [ISBN: 0-444-86319-2] Introduces holomorphic functions defined on domains in (infinite-dimensional) locally convex spaces. Emphasizes properties of the various locally convex topologies which spaces of such functions admit. Hundreds of exercises, many with notes. With appendices, the development is self-contained. PZ

Differential Equations, P*. The Numerical Solution of Nonlinear Problems. Ed: Christopher T.H. Baker, Chris Phillips. Clarendon Pr, 1981, viii + 369 pp, $39. [ISBN: 0-19-853354-3] Fifth volume in a sequence of proceedings of annual summer school lectures at Liverpool University. On numerical solutions of initial-value and boundary-value problems in ordinary differential equations, of partial differential equations and of integral equations, plus chapters on approximation and on programming techniques. Extensive reference list. JK

Numerical Analysis, S(15-18), P, L*. Handbook of Applicable Mathematics, Volume III: Numerical Methods. Ed: Robert F. Churchhouse. Wiley, 1981, xvii + 565 pp, $85. [ISBN: 0-471-27947] Aimed at non-mathematicians who have some mathematical sophistication. Includes chapters on systems of linear equations, interpolation, curve fitting, non-linear equations, matrix computations, quadrature, ordinary differential equations, partial differential equations, integral equations, optimization, and an appendix of Fortran programs. Index. RJA

Numerical Analysis, S(15-16), L*. An Introduction to the Approximation of Functions. Theodore J. Rivlin. Dover, 1981, viii + 150 pp, $3.50 (P). [ISBN: 0-486-64069-8] Uniform, least squares and least-first-power approximations. Polynomials, splines and rational functions. (Original edition, TR, October 1969.) RWN

Numerical Analysis, P. Approximation on a Rectangular Grid with Application to Finite Element Methods and Other Problems. S.G. Mikhlin. Trans. and Ed.: R.S. Andersen, T.O. Shaposhnikova. Sijthoff & Noordhoff, 1979, xi + 224 pp, $35. [ISBN: 90-286-0008-6] Research monograph on the constructive theory of basis functions for the finite element method. Includes completeness, best order of approximation and stability. Applications. Presumes functional analysis. RWN

Functional Analysis, P. Functional Analysis and Approximation. Ed: P.L. Butzer, B. Sz.-Nagy, E. Görlich. ISNM, No. 60. Birkhauser Boston, 1981, 482 pp, $47.95. [ISBN: 3-7643-1212-2] Proceedings of an August 1980 international conference at Oberwolfach; includes memorials to Lionel Cooper, a special bibliography on Bernstein polynomials, as well as a list of 22 new and unsolved problems. LAS

Analysis, T(16: 1), S, P, L. Kronecker Products and Matrix Calculus: With Applications. Alexander Graham. Math. & its Appl. Ellis Horwood, 1981, 130 pp, $39.95. [ISBN: 0-85312-391-8] An introduction to Kronecker (tensor) matrix products which is rigorous and starts from first principles. Includes some applications, lots of examples, a bibliography and problems. CEC

Differential Geometry, P. Manifolds and Lie Groups: Papers in Honor of Yozô Matsushima. J. Hano, et al. Progress in Math., V. 14. Birkhauser Boston, 1981, xii + 459 pp, $30. [ISBN: 3-7643-3053-8] A collection of some 24 papers written for the celebration of the 60th birthday of Yozô Matsushima; all dealing with topics in the general area of manifolds and Lie groups. JS

Differential Geometry, P. The Theory of Eisenstein Systems. M. Scott Osborne, Garth Warner. Pure and Appl. Math., V. 99. Academic Pr, 1981, xiii + 385 pp, $55. [ISBN: 0-12-529250-3] An attempt to lay the foundations for the theory and stimulate interest in the development of a trace formula for the nonuniform case extending that which Selberg developed for a reductive Lie group G with a uniform lattice Γ in G. Assumes familiarity with reductive Lie groups, their lattices, and automorphic forms. References, indexes. JS

Algebraic Topology, T(16-18: 1, 2), S, P. Commutator Calculus and Groups of Homotopy Classes. Hans Joachim Baues. London Math. Soc. Lect. Note Ser., V. 50. Cambridge U Pr, 1981, 160 pp, $23.95 (P). [ISBN: 0-521-28424-4] The first part encompasses homotopy operations, nilpotent group theory and nilpotent Lie algebra theory, while part two discusses homotopy theory over a subring of the

rationals containing 1/2 and 1/3. Bibliography. Index. RJA

Topology, S(18), P. Uniform Structures on Topological Groups and their Quotients. W. Roelcke, S. Dierolf. McGraw-Hill, 1981, xi + 276 pp, $44.95. [ISBN: 0-07-053412-8] A systematic and thorough treatment of the four natural uniformities on a topological group (left, right, supremum, and infinimum) dealing with questions related to subgroups, product groups, representations, quotients, and completeness. Exercises, index, bibliography. JS

Operations Research, T(15: 1), L. Linear Programming and Applications, A Course Text. Will McLewin. Input-Output Pub, 1981, xiv + 216 pp, (P). [ISBN: 0-904870-12-X] First half develops simplex method, duality and revised simplex methods using both tableau and matrix operations approaches. Remainder covers various applications including Shor-Khachian ellipsoid method, network flows, and game theory. Few exercises. JRG

Optimization, P. The Theory of Subgradients and its Applications to Problems of Optimization, Convex and Nonconvex Functions. R.T. Rockafellar. Res. and Educ. in Math., No. 1. Heldermann Verlag, 1981, vii + 107 pp, 28 DM (P). [ISBN: 3-88538-201-6] New developments in theory of subgradients of functions on R^n, and how non-differentiable functions arise naturally in optimization problems. Relationship of subgradients to dual pairs of problems; monotonicity questions for subgradient multi-functions. JRG

Optimization, P*, L. Linear and Combinatorial Optimization in Ordered Algebraic Structures. U. Zimmermann. Annals of Discrete Math., No. 10. Elsevier North Holland, 1981, ix + 380 pp, $61. [ISBN: 0-444-86153-X] Results and methods for linear and combinatorial optimization problems where linear functions over the reals (or integers) are replaced by functions linear over algebraic structures, e.g., ordered semi-modules. Part 1 develops theory of ordered algebraic structures; Part 2 discusses optimization problems over these structures, including path, eigenvalue, and network flow problems. Extensive bibliography. JRG

Optimization, T(16-17: 1). Optimality in Parametric Systems. Thomas L. Vincent, Walter J. Grantham. Wiley, 1981, xv + 243 pp, $29. [ISBN: 0-471-08307-0] Unified theoretical approach to parametric optimization for static and dynamic systems. Discusses parametric game theory in nonlinear programming setting. JRG

Probability, P. Characterization of Optimal Strategies in Dynamic Games. L.P.J. Groenewegen. Math. Centre Tracts, No. 90. Math Centrum, 1981, 110 pp, Dfl. 14,70 (P). [ISBN: 90-6196-156-4] Characterization of optimality for fairly general decision processes: discrete and continuous processes on general state and action spaces. Hence proofs are independent of the specific structure of the process. JRG

Statistics, T(17-18: 1, 2), S, P. Sequential Nonparametrics: Invariance Principles and Statistical Inference. Pranab Kumar Sen. Wiley, 1981, xv + 421 pp, $42.50. [ISBN: 0-471-06013-5] Lays the foundation of a theory of nonparametric sequential inferences. Develops various types of robust sequential procedures. Presupposes familiarity with parametric sequential analysis, the usual nonparametric methods, and abstract probability theory. FLW

Statistics, T(16-18: 1), S, P, L*. Discrimination and Classification. D.J. Hand. Wiley, 1981, x + 218 pp, $36.95. [ISBN: 0-471-28048-8] Treats statistical discriminant analysis, pattern recognition, and cluster analysis. The variable selection problem and both continuous and categorical variables are considered. FLW

Statistics, T(15-17: 1), S, L. Statistical Methods for Rates and Proportions, Second Edition. Joseph L. Fleiss. Wiley, 1981, xviii + 321 pp, $28.95. [ISBN: 0-471-06428-9] Four-fold tables, differences in proportions, types of studies, comparing several proportions, misclassification errors, interrater agreement. Presupposes prior statistical study but little mathematics. A few exercises are supplied. Most of the examples are from the health sciences. FLW

Statistics, P*. Interpreting Multivariate Data. Ed: Vic Barnett. Wiley, 1981, xvi + 374 pp, $54.95. [ISBN: 0-471-28039-9] In the Wiley Series in Probability and Mathematical Statistics. Proceedings of the Conference entitled "Looking at Multivariate Data" held at the University of Sheffield, U.K., in March 1980. Contains expanded versions of the invited papers which review existing methods in multivariate exploratory data analysis and illustrate their application in many fields. Extensive references. RSK

Statistics, P. Current Topics in Survey Sampling. Ed: D. Krewski, R. Platek, J.N.K. Rao. Academic Pr, 1981, xv + 509 pp, $29.50. [ISBN: 0-12-426280-5] Proceedings of the International Symposium on Survey Sampling held at Carleton University in Ottawa in May 1980. Contains revised versions of all invited papers, abstracts of contributed papers, and edited versions of the discussions following the invited presentations. Topics covered include nonsampling errors, current survey research activity, superpopulation models, variance estimation, and imputation techniques. RSK

Statistics, P. Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface. Ed: William F. Eddy. Springer-Verlag, 1981, xv + 378 pp, $24 (P). [ISBN: 0-387-90633-9] The papers included in these proceedings cover a wide range of issues related to the use of computers in statistical research. AO

Statistics, P. Classification et Analyse Ordinale des Données. I.C. Lerman. Dunod, 1981, 740 pp, 140 FF (P). [ISBN: 2-04-015405-1] A comprehensive survey of new methods of data classification and ordinal analysis appropriate for the computer age: statistical, algebraic and combinatorial techniques representing quantitative, qualitative and contingency data. Part I provides a theoretical synthesis of methods; Part II presents case studies (in areas such as geography, medicine and archaeology) illustrating and applying the general methods. LAS

Computer Literacy, T(13: 1), S, L. Understanding Computer Systems. Harold W. Lawson, Jr. Computer Sci Pr, 1982, vii + 164 pp, $15.95 (P). [ISBN: 0-914894-31-5] An impressive explanation of computer systems and vocabulary for blue-collar workers using as an extended metaphor washing and drying dishes: dishracks become buffers; a crying baby is a high priority interrupt. This no-holds-barred treatment of everything from parity bits to hand gates is motivated by Scandinavian unions' labor agreements that workers using computer equipment have a right to "gain insight into the fundamental features" of computer systems by instruction in "language easily understood by persons lacking special knowledge." LAS

Computer Programming, T(13-18: 1), S*. Programmer's Guide to LISP. Ken Traction. TAB Books, 1980, 210 pp, $6.95 (P). [ISBN: 0-8306-1045-6] Part one uses a question and answer approach to present the LISP language, its syntax, semantics, and dialects. Part two contains many examples of LISP programs. Informal, attractive style of organization. Index. RJA

Computer Programming, T(13-18: 1), S. The ADA Programming Language: A Guide for Programmers. I.C. Pyle. Prentice-Hall, 1981, x + 293 pp, $14.95 (P). [ISBN: 0-13-003921-7] Written for practicing programmers of embedded or real-time computer systems. Presumes a knowledge of some other high level programming language(s). Comprehensive treatment. Begins with basic ADA features, followed by chapters on exceptions, packages, and parallel programming. Later part deals with program structure and machine specific issues. Chapter exercises. Several appendices. RJA

Computer Programming, T(13-18: 1). Assembly Language for the PDP-11. Charles Kapps, Robert L. Stafford. Prindle, Weber & Schmidt, 1981, xiv + 353 pp. [ISBN: 0-87150-304-2] Presumes some programming experience in a high-level programming language. Thorough discussion of machine organization. Integrates theory and practice. Complete presentation of all aspects of machine and assembly level programming on the PDP-11. Last chapter provides a good introduction to operating systems and systems programming. Exercises. Several useful appendixes on running programs and many subsystems. Glossary. Index. Summaries of instructions inside book covers. RJA

Computer Programming, S. 6502 Assembly Language Subroutines. Lance A. Leventhal, Winthrop Saville. Osborne/McGraw-Hill, 1982, x + 550 pp, $12.99 (P). [ISBN: 0-931988-59-4] The first three chapters of this book provide a basic introduction to the architecture and assembly language of the 6502 microprocessor. The remaining chapters are a collection of assembly language subroutines for performing a variety of common tasks. AO

Software Systems, T(15-16), S*, P*, L. Software Tools in Pascal. Brian W. Kernighan, P.J. Plauger. Addison-Welsey, 1981, xi + 366 pp, $15.95 (P). [ISBN: 0-201-10342-7] A revision of the 1966 classic offering Pascal code for the basic tools of a UNIX-like computing environment: sorting, filters, files, editing, formatting macroprocessing. The general theme is to control complexity by writing code in simple, cohesive modules. (The software for the book is available for $45 from the publisher.) LAS

Software Systems, T(14-16: 1), S, L. A Practical Introduction to Computer Graphics. Ian O. Angell. Halsted Pr, 1981, ix + 146 pp, $16.95 (P). [ISBN: 0-470-27251-1] A very down-to-earth treatment, replete with increasingly, complex sample programs, beginning with squares and circles, moving through two and three dimensional transformations to hidden line algorithms and computer movies. An excellent blend of elementary mathematics (primarily matrix transformations) and interesting computer methods. LAS

Computer Science, T*(15-18: 1, 2), S, L. Formal Specification of Programming Languages: A Panoramic Primer. Frank G. Pagan. Prentice-Hall, 1981, x + 245 pp, $19.95. [ISBN: 0-13-329052-2] Provides an informal survey of several metalanguages for the formal specification of programming languages. Includes chapters on formal syntax and formal semantics and the transition between the two, and on programming languages themselves used as metalanguages. Exercises. Bibliography. Index. RJA

Computer Science, T*(14-18: 1, 2), S, L. Mathematical Foundations of Programming. Frank S. Beckman. Addison-Wesley, 1980, xviii + 443 pp, $19.95. [ISBN: 0-201-14462-X] Descriptive presentation of the mathematics underlying computer programming and computer design. Includes chapters on effective procedures, foundations of mathematics, recursive functions, Turing machines, computability, automata, formal languages, computational complexity. Chapter exercises and references. Index. RJA

Computer Science, P. Program Flow Analysis: Theory and Applications. Steven S. Muchnick, Neil D. Jones. Prentice-Hall, 1981, xvii + 418 pp, $23.95. [ISBN: 0-13-729681-9] Flow analysis attempts to discover properties of the run-time behavior of a program without actually running it. Text divides into four parts: techniques and availability; applications to program optimization; applications to software engineering; theoretical cosiderations. Bibliography. Subject index. Name index. RJA

Computer Science, T(13-15: 1). Introduction to Microcomputing. Sydney B. Newell. Harper & Row Pub, 1982, xii + 615 pp, $24.95. [ISBN: 0-06-044802-4] Designed for an introductory course on

microcomputer organization and assembly language programming, this text deals exclusively with the Motorola 6800 microprocessor. The traditional topics are covered. An adequate number of exercises is provided at the end of each chapter. AO

Systems Theory, T(18), P. Nonlinear System Theory: The Volterra/Wiener Approach. Wilson J. Rugh. Johns Hopkins U Pr, 1981, xiv + 325 pp, $32.50. [ISBN: 0-8018-2549-0] Represents a middle ground between relatively simple techniques (of limited generality) and more sophisticated techniq... (of limited applicability). Considers systems composed of feedback-free interconnections of linear dynamic systems and simple static nonlinear elements. JRG

Applications (Artificial Intelligence), P. The Handbook of Artificial Intelligence, Volume I. Ed: Avron Barr, Edward A. Feigenbaum. William Kaufmann, 1981, xiv + 409 pp, $30. [ISBN: 0-86576-005-5] Present volume contains contributions to several areas of current research in artificial intelligence: search methods and programs; knowledge representation; understanding natural language; understanding spoken language. Bibliography, name index, subject index for Volume I. RJA

Applications (Artificial Intelligence), P. Knowledge-Based Systems in Artificial Intelligence. Randall Davis, Douglas B. Lenat. McGraw-Hill, 1982, xxi + 490 pp, $39.50. [ISBN: 0-07-001557-7] Reports on two case studies in knowledge engineering: (1) AM which deals with paradigms of mathematical discovery and scientific research; (2) TIERESIAS which examines the problem of building and using knowledge-based systems. RJA

Applications (Artificial Intelligence), P, L. The Brains of Men and Machines. Ernest W. Kent. BYTE/McGraw-Hill, 1981, ix + 286 pp, $15.95. [ISBN: 0-07-034123-0] A comparative study of brain physiology and computer architecture, relating both to parallel trends in evolution, in function, and in goals. LAS

Applications (Biology), P, L. Mathematical Models of Epidemics. H.A. Lauwerier. Math. Centre Tracts, V. 138. Math Centrum, 1981, iii + 121 pp, Dfl. 15,75 (P). [ISBN: 90-6196-216-1] Analysis of the Kermack-McKendrick model, both the long-neglected original, and a simplified version revived by Kendall (1970). Confirms logistic function as standard way of describing epidemic behavior. JRG

Applications (Biology), P. Lecture Notes in Biomathematics-43: Mathematical Modeling of the Hearing Process. Eds: Mark H. Holmes, Lester A. Rubenfeld. Springer-Verlag, 1981, 104 pp, $9.80 (P). [ISBN: 0-387-11155-7] Proceedings of a July 1980 NSF-CBMS regional conference held at Rensselaer Polytechnic Institute in Troy, New York. Six papers extending traditional models to account for recent experimental evidence. LAS

Applications (Biology), S(16-17), P. Differential Equations and Applications in Ecology, Epidemics, and Population Problems. Ed: Stavros N. Busenberg, Kenneth L. Cooke. Academic Pr, 1981, xv + 359 pp, $34.50. [ISBN: 0-12-148360-6] The proceedings of a conference held at Harvey Mudd College during January 1981. The papers treat a variety of questions in population dynamics as well as a number of topics in differential and functional differential equations with applications in mathematical biology. AO

Applications (Computer Graphics), T(15-17: 1), L*. Algorithms for Graphics and Image Processing. Theo Pavlidis. Computer Sci Pr, 1982, xvii + 416 pp, $24.95. [ISBN: 0-914894-65-X] This text emphasizes the mathematical tools used in pictorial information processing by computers (computer graphics, image processing, and pictorial pattern recognition). Parts of the book presume a background which includes calculus, elementary statistics, elementary graph theory, linear algebra, geometry, data structures, and programming. AO

Applications (Economics), S(15-17), P, L.** Lecture Notes in Economics and Mathematical Systems-194: Cost and Production Functions. Ronald W. Shephard. Springer-Verlag, 1981, xi + 104 pp, $12 (P). [ISBN: 0-387-11158-1] Reprint of 1953 original Princeton University Press classic, in which convex analysis is used to demonstrate the duality between cost and production functions. "One of the most original contributions to economic theory of all time"--from the foreward by Dale Jorgensen. A beautiful application of advanced calculus, readily accessible to advanced undergraduate mathematics students. LAS

Applications (Geophysics), P. The Solution of the Inverse Problem in Geophysical Interpretation. Ed: R. Cassinis. Ettore Majorana Intern. Sci. Ser., V. 11. Plenum Pr, 1981, ix + 381 pp, $45. [ISBN: 0-306-40735-3] To a geophysicist, the inverse problem is quite general: to determine features of a concealed source by observing its effects. This volume contains lectures and contributed papers on diverse aspects of this problem from the Third Course of the 1980 meeting of the International School of Applied Geopysics. LAS

Applications (Pattern Recognition), T(15-18: 1, 2), S, P. Pattern Recognition with Fuzzy Objective Function Algorithms. James C. Bezdek. Plenum Pr, 1981, xv + 256 pp, $35. [ISBN: 0-306-40671-3] Text organized around fuzzy relations and partitions. Includes chapters on models for pattern recognition, objective function clustering, cluster validity, modified objective function algorithms, applications in classifier design. Exercises. References. Algorithm index. Author index. Subject index. RJA

Applications (Pattern Theory), P. Regular Structures: Lectures in Pattern Theory, Volume III. Ulf Grenander. Appl. Math. Sci., V. 3. Springer-Verlag, 1981, viii + 569 pp, $24 (P). [ISBN: 0-387-

90560-X] Includes chapters on basic principles, regularity, metric pattern theory, topological image algebras, pattern syntheses, taxonomic patterns, patterns in mathematical semantics. Appendix. Notes. Bibliography. Index. RJA

Applications (Physics), T(17-18: 1, 2), S*, P. Gauge Theory and Variational Principles. David Bleecker. Global Analysis, Pure & Appl. Addison-Wesley, 1981, xviii + 179 pp, $17.50. [ISBN: 0-201-10096-7] A careful but sophisticated introduction which builds the geometric machinery necessary to give a clear picture of some of the recent geometric developments in physics. No exercises, good references. Requires some significant experience with differential geometry. JAS

Applications (Physics), P*. The Racah-Wigner Algebra in Quantum Theory. Lawrence C. Biedenharn, James D. Louck. Ency. of Math. & its Appl., V. 9. Addison-Wesley, 1981, lxxxviii + 534 pp, $54.50. [ISBN: 0-201-13508-6] This monograph is a sequel to Angular Momentum in Quantum Physics: Theory and Application by the same authors. It develops the theory of angular momentum within the context of the algebra of bounded operators acting in a Hilbert space. AO

Applications (Physics), P. Lagrangian Analysis and Quantum Mechanics: A Mathematical Structure Related to Asymptotic Expansions and the Maslov Index. Jean Leray. Transl: Carolyn Schroeder. MIT Pr, 1981, xvii + 271 pp, $35. [ISBN: 0-262-12087-9] The author suggests that an alternate title for this work might have been The Introduction of Planck's Constant into Mathematics. It is an account of recent mathematical developments arising from the study of asymptotic solutions to differential equations in which quantum conditions arise in a natural way. AO

Applications (Simulation), P. The 13th Annual Simulation Symposium. Ed: Victor P. Boyd, et al. Annual Simulation Symposium (P.O. Box 22621, Tampa, FL 33622), 1980, viii + 337 pp, (P).

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Southeastern Section

The sixty-first annual meeting of the Southeastern Section was held on April 9-10, 1982 at Emory University, Atlanta, Georgia. A total of 294 persons attended the meeting, including 248 members of the Association and 35 students.

Invited Lectures:

"Points are Simpler Than Circles," by C. Ray Wylie, Greenville, South Carolina.
"Mathematics Tomorrow," by Lynn Arthur Steen, St. Olaf College.
"Error-correcting Codes and Some Applications," by Marshall Hall, Emory University and California Institute of Technology.

One Hour Lectures:

"What To Do With Talented Undergraduates," by Elwood G. Parker, Guilford College.
"Barnaby and the Purple Heptagon," by J.B. Stroud, Davidson College.

Short Presentations:

"Let's Really Use the Computer to Teach Calculus," by Linda Boyd and Charles Stone, DeKalb Community College.
"A Simple Theorem on the Asymptotics of Riemann Integration, Based on Classroom Experience," by Lance D. Drager, Georgia Institute of Technology.
"The Mathematics Recommendation in the College Board's Project Equality," by John Kenelly, Clemson University.
"Writing in Mathematics Classes," by Beth Hardy, Georgia Southern College.
"MAWIS (Mathematics At Work In Society)," by John Neff, Georgia Institute of Technology.

2. P. R. Halmos, E. E. Moise, and G. Piranian, The problem of learning to teach, this MONTHLY, 82 (1975) 466–476.

3. F. Burton Jones, The Moore method, this MONTHLY, 84 (1977) 273–8.

4. J. A. Murtha, The apprentice system for math methods, this MONTHLY, 84 (1977) 473–6.

5. Warren Page, A small group strategy for enhanced learning, this MONTHLY, 86 (1979) 856–8.

6. J. Weinglass, Small groups: an alternative to the lecture method, Two-year Coll. Math. J., 7 (1976) 15–20.

HOW TO CONSTRUCT AND ANALYZE PROOFS—A SEMINAR COURSE

ROBERT B. REISEL

Department of Mathematical Sciences, Loyola University, Chicago IL 60626

In order to learn mathematics, it is necessary to be able to understand proofs. Mathematics students or graduates who do not know how to construct proofs or how to read them critically have, at best, an incomplete understanding of any mathematics they are studying. A major goal in teaching mathematics is, therefore, to teach proofs and proof techniques. The corresponding problem is when and how to do this.

Traditionally, students first studied proofs in high school geometry courses. They still do, although now proofs are also found in high school algebra courses. The trouble is that rigorous proofs are slighted (and rightly so, I believe) in the usual college calculus courses. By the time students study abstract mathematics, they have forgotten whatever they did know about proofs and are at a loss when asked to write or to analyze a proof. Instructors must take time from the actual mathematical content of their courses to teach proof construction and criticism. With the increase in the amount of mathematics and computer science in the curriculum, however, it is difficult to justify taking time to teach techniques of theorem-proving by omitting important topics.

At Loyola University we decided to solve this problem by running a seminar for sophomore mathematics and computer science majors which would have as its goal the teaching of proof construction and criticism. In the sophomore year these students have a reasonable degree of mathematical maturity from having finished our calculus sequence, but they have not begun to study abstract algebra or real analysis. We decided on a seminar rather than on a more traditional course so that the students will get actual practice in constructing and analyzing proofs. Individual assignments of proofs to be constructed are given to the students who then present their results orally before the seminar. Since it is to be expected that some of the proofs will be incorrect or incomplete, the other students in the seminar must be alert to raise questions and objections. This helps develop a critical sense in the students more effectively than the situation in which students only raise questions about correct proofs given in their texts or by their instructors. An extra bonus of the seminar approach is that it gives the students the experience of learning to think on their feet by trying to answer questions or make corrections on the spot.

We were next faced with the problem of what topic should be studied in the seminar. It should be one that the students recognize as something important and useful. In fact, it should be one that the students find so interesting that they will not only be thinking about the goal of learning to construct proofs, but also will be studying the subject for its own sake. The topic should not require much prior knowledge or experience, the concepts should be clear, and the proofs should be relatively short and straightforward. We finally chose the elementary parts of the theory of metric spaces. Not only does this subject have the above attributes, but also it leads directly from calculus to more rigorous courses in analysis and topology.

Since we were not able to find a book that met our needs, we prepared a set of seminar notes (see [**1**]). Although other topics could be chosen, I will give a brief description of these notes so that the general level and organization of the seminar can be seen. The first chapter contains a brief explanation of those parts of logic that are needed in proofs. This is not a theoretical

discussion, but it is restricted to an informal description of the logical connectives and quantifiers and an explanation of the structure of direct proofs, indirect proofs and proofs by *reductio ad absurdum*. The students are asked to try to construct some simple proofs and compare their efforts with the correctly written proofs included in the notes. The students should read and work through this material before attempting any actual assignments. There is next a chapter on sets and mappings, emphasizing inverse and direct images, injections, surjections, bijections, and inverse mappings. Definitions and theorems are stated, along with brief explanations. Hints for the proofs are included, but the proofs are not, except for some that illustrate techniques that might be unfamiliar to the students. The following chapter covers the geometry of metric spaces, e.g., open and closed sets. The next chapter concerns continuity of mappings and leads from the familiar ideas of calculus to topology, although topological spaces are not formally presented. The remaining chapters, which are independent of each other and are not usually covered completely, take up sequences and completeness, connectivity and compactness, with an emphasis on the metric space of real numbers.

One difficulty students have with proofs comes from the different styles used in high school and in college texts. The former uses a formal presentation with the "given" and the "to prove" carefully written, the steps of the proof numbered, and complete reasons written for each step. On the other hand, college texts adopt a narrative style for proofs in which easily supplied steps are omitted. This latter style is the one used in most mathematical writing and students must adjust to it. The high school style has the advantage of training students to write careful and complete proofs which they can easily examine critically. We have the students start, therefore, by writing their proofs in this style, and then we gradually loosen the formality until they are using the narrative form. The notes illustrate this method by giving examples of some of the proofs, following this progression of styles.

We limit each section of the seminar to at most ten students. The seminar is a two-semester-hour course with two fifty-minute meetings a week. The first week is spent in organizing the seminar, making assignments, and helping the students understand the initial material on the logic of proofs. The first assignments are short, usually just two short proofs each. This is done so that the talks can start in the second week and everyone will make a presentation in the first few weeks. Later assignments are longer; we usually try to make them long enough so that at most two reports can be given in a meeting. The students not making reports are, of course, taking notes as they would in a regular course, questioning the speaker about unclear points, and observing the good and bad features of the presentation. This helps them develop a critical sense about proofs and improve their own talks.

The instructor is available to help the students prepare their presentations. We ask the students to write out their work completely and show it to the instructor before their talks so that errors can be caught and stylistic suggestions made. The students are asked not to look up proofs in other books, but to consult with their instructor if they have problems. As the semester proceeds, we encourage the students to rely more on their own work and less on the instructor so that they build up self-confidence. The instructor attends all the seminar meetings to observe the presentations. He does not make corrections unless none of the class has noted the error. Since he observes the progress of the students, it is not necessary to have examinations. Some of the instructors give a written assignment also to the entire class to see if the students have learned written, as distinguished from oral, techniques.

We are now in the second year of operation of the seminar. We are pleased with the results and the students' reactions have been positive. Many students have expressed enthusiasm about the seminar, pointing out how it has given them an understanding of proofs and a self-confidence both in constructing proofs and in presenting them before the class. They seem to have become familiar with the concepts and techniques of metric spaces; this should certainly help them in future analysis courses. It is still too early to compare students who have completed the seminar with those who have not, but we are confident that the ones who have will show greater facility in

doing abstract mathematics. In fact, we have made the seminar a prerequisite for abstract algebra and real analysis courses.

In conclusion, we believe that a seminar which teaches proof construction and analysis fills an important need in the mathematics curriculum and is an effective way of reaching these goals.

Reference

1. R. B. Reisel, Elementary Theory of Metric Spaces. A Course in Constructing Mathematical Proofs, Springer-Verlag, New York, 1982.

ACTUARIAL SCIENCE

BETTE WARREN
Department of Mathematical Sciences, SUNY at Binghamton, NY 13901

The Department of Mathematical Sciences at the State University of New York (SUNY) at Binghamton has been involved in actuarial training for the past five years, developing a program which is inexpensive to operate and complements the traditional curriculum. Since there seems to be growing interest among college faculty members in applied programs in general and actuarial science in particular, we thought it appropriate to publish this outline of our operation.

The Actuarial Examinations. Full admission into the actuarial profession is based in part upon a series of difficult examinations. In the United States, a life actuary must pass ten examinations (referred to as "parts 1–10") sponsored by the Society of Actuaries and a casualty actuary must pass ten (also referred to as "parts") sponsored by the Casualty Actuarial Society. An interested student has ample time to decide which of these two paths to follow because the first four examinations are the same and are administered jointly by the two societies. It is common for prospective actuaries to pass from one to three parts prior to graduation from college and to take the remainder while employed. A college (undergraduate) mathematics or statistics department can justify conducting a program to prepare students for the first three or four examinations: few students progress beyond this point while still undergraduates, and these early examinations are heavily mathematical.

Our Program at SUNY at Binghamton. Our actuarial students are drawn almost exclusively from the department's 200 or so undergraduate majors. They are expected to fulfill all normal degree requirements and are not segregated; the formal exam preparation (for parts 1–3) carries few credits and cannot be used to fulfill major requirements, so it must be pursued in addition to rather than instead of taking standard mathematics courses. Since the material required for parts 1 and 2 is covered in the regular curriculum, students begin their actuarial training by taking prescribed mathematics courses. They are then encouraged to enroll in one-credit problem solving seminars. Preparation for parts 3 and 4 is handled differently because the material on the syllabus is not fully covered in our standard courses and there are usually few students involved at any one time.

Students taking the part 1 seminar are expected to have completed three semesters of calculus (including multivariate calculus) and one of linear algebra. Part 2 students take a two-semester sequence in probability and mathematical statistics prior to attending their seminar. Each of these seminars meets once each week for 85 minutes, and carries one credit. While there is some review of core material, most of the class time is devoted to analyzing and explaining solutions to problems prepared outside of class by students. The instructor comments frequently on the students' solutions, offering different modes of attack and emphasizing the general principles involved. It is quite important that students learn not simply how to do the problems, but how to do them efficiently. We use problems from practice examinations supplied by the actuarial

societies, instructor generated problem sets, and commercial sources such as Shaum's Outlines. The best of these sources are the practice examinations because they are similar in form (multiple choice), content, and phrasing to the actual examinations. A part 1 study guide and problem manual has just been published by the Graduate School of Actuarial Science at Northeastern University (distributed by ACTEX, P. O. Box 2392, Framingham, MA 01701). Although it is too early to make any comments on the use of these new part 1 materials, we can say that Northeastern's manuals for later examinations have proven quite valuable.

We encourage broad participation in both of these seminars, believing that the students (the vast majority of whom will not become actuaries) all benefit by sharpening their basic mathematical and technical skills and increasing their sophistication in solving problems.

Numerical analysis and operations research are the subjects currently covered on the third examination. Because our courses in these areas differ markedly from the examination material, we have found it necessary to have the third examination seminar meet twice as often as the other two. Most of the semester is spent working through the required texts *Fundamentals of Numerical Analysis* by S. Kellison (published by Irwin, 1975) and *A Study Manual for Operations Research* edited by E. Narragon (published by the Society of Actuaries and the Casualty Actuarial Society, 1980), adhering to a schedule which leaves the last few weeks open for review and extra problems. The Northeastern manuals are helpful because they contain both end-of-chapter review problems which are useful as new material is being presented, and mixed multiple choice tests which are helpful during the final review. The actuarial societies also supply practice examinations. Within the past year this seminar has attracted a couple of actuary trainees from local insurance companies, and we have had to become more flexible in scheduling class meetings than when we were dealing only with traditional students.

The fourth examination tests the theory of interest and life contingencies, i.e., the valuation of various payment streams taking into consideration both interest rates and relevant probabilities, such as the probability of death or disability. The required texts are *The Theory of Interest* by S. Kellison (published by Irwin, 1970) and *Life Contingencies* by C. W. Jordan (published by the Society of Actuaries, 1975). While a course covering these topics might be desirable, we have not had enough students get this far to justify offering it. A student studying for part 4 spends roughly a semester and a half working through the two books, and the remainder of the final semester solving problems in sources such as the Northeastern manual. The credit granted is not fixed since this is handled informally through independent study.

The examinations are given twice yearly in mid-November and mid-May. Because the November date is in the middle of our fall semester, we have offered the seminars only in the spring. This does cause problems. A student completing the relevant course work must either take the examination in November without benefit of the seminar, wait until the following May, or try to take the seminar concurrently with the course work. The obvious alternative is to offer seminars in the fall and have them meet twice as often for the half semester preceding the examination. We are going to try this next fall.

A Few Statistics. As was mentioned previously, we encourage all students with the proper course background to try the first two seminars, and consequently the initial enrollments are substantial; normally between 40 and 50 students sign up for the part 1 seminar and 15 to 25 for the part 2 seminar. The drop rate, however, is also substantial, generally in the range of 30 to 50 percent. About 50 percent of the students who take one of the first two examinations immediately after completing the seminar pass, and of those failing once and taking it a second time, about two-thirds pass. While the drop rate seems high and pass rate low when comparing them to drop and pass rates in a normal course, it must be remembered that these students are trying to meet an uncompromising external standard. The level of conceptual understanding and technical competence implied by a passing mark on an actuarial examination should be equated with that implied by a high grade, not simply a passing grade, in a normal course, and the student who lacks

the self-discipline to study adequately cannot "squeak by." The pass rates compare favorably with the national rate, which is just under 40% (not differentiating between first-timers and others), while the pass rate for our students who attempt the examinations without first taking the seminar is lower than the national rate. The figures are about the same for the third seminar, except that both the drop rate and the initial enrollment are small.

In the last four years we have had approximately 70 students pass part 1 and 25 pass part 2. Four have passed part 3 and one part 4, but these are within the last two years. More than 20 seminar students have taken actuarial jobs with major insurance companies, and several others declined offers. Everyone who has passed at least one examination and applied for actuarial positions has received several offers.

Information Available. More detailed information such as examination syllabi and syllabus revisions, lists of texts and publishers, and dates, locations and fee schedules for examinations can be obtained from the Society of Actuaries (208 South La Salle Street, Chicago, IL 60604), and the Casualty Actuarial Society (One Penn Plaza, New York, NY 10001). The societies can also supply the address of the actuarial club nearest you, a list of companies employing Society of Actuaries members, and a list of colleges and universities offering actuarial science programs. The actuarial clubs usually maintain lists of actuaries who would be willing to describe and promote the profession to a group of students. The list of actuarial programs is valuable not only to the organizer of a new program, but also to students who may be interested in graduate programs in actuarial science. While the information mentioned thus far is available without cost, the most convenient employer listings appear in the Society of Actuaries' year book which costs $50. This listing is very helpful to students applying for jobs, especially jobs in distant sections of the country.

There is also a series of nine examinations sponsored by the American Society of Pension Actuaries. However, since our resources are limited and these tests are narrower in scope and seem to offer less career flexibility than those of the life and casualty societies, we do not prepare students for them. If you would like more information, write to the society at 1700 K Street NW, Washington, DC 20006.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Product of Reflection Matrices, $a_i^* H a_i = 0$

S 24 [1980, 60]. *Proposed by Dragomir Z. Djokovic and Jerry Malzan, University of Waterloo, Waterloo, Ontario, Canada.*

If u is a unit column vector in C^n, let $R(u)$ be the reflection matrix $I_n - 2uu^*$. Let $a_1, \dots, a_n$ be a basis for C^n consisting of unit vectors and $R_i = R(a_i)$. Define $A = R_1 R_2 \cdots R_n$ and $H = (A + I_n)(A - I_n)^{-1}$. [Note that $A - I_n$ is invertible.] Show that $a_i^* H a_i = 0$ for $i = 1, \dots, n$.

Solution by R. K. S. Rathore and C. S. K. Chetty, Indian Institute of Technology, Kanpur, India. First, let us give a proof for the invertibility of $A - I_n$. Let, on the contrary, $Ax = x$ for some nonzero vector x. As $R_1^{-1} = R_1^* = R_1$, $R_2 R_3 \cdots R_n x = R_1 x$. Hence for some scalars α_j,

$$x + \sum_{j=2}^{n} \alpha_j a_j = x - 2 a_1 a_1^* x.$$

Cancelling x, by linear independence of a_j's it follows that $a_1^* x = 0$. Hence $R_1 x = x$ and $R_3 R_4 \cdots R_n x = R_2 x$. Continuing the above process, we get $0 = a_1^* x = a_2^* x = \cdots = a_n^* x$. As a_j's span C^n, we arrive at a contradiction, namely, $x = 0$ which establishes the invertibility of $A - I_n$.

To solve the main problem, let $i \in \{1, 2, \dots, n\}$ and $(A - I_n)x = a_i$. Write $R = R_{i+1} \cdots R_n$ and $L = R_1 \cdots R_{i-1}$. Then $Rx - 2a_i a_i^* Rx = R_i Rx = L^{-1} Ax = R_{i-1} \cdots R_1 (a_i + x)$ which may also be rewritten as

$$x + \sum_{j=i+1}^{n} \beta_j a_j - 2 a_i a_i^* Rx = a_i + x + \sum_{j=1}^{i-1} \beta_j a_j,$$

β_j's being some scalars. Again cancelling x and using linear independence of a_j's, it follows that $Rx = x$ (since all β_j's are zero) and $-2a_i^* Rx = 1$ and hence $-2a_i^* x = 1$. Therefore

$$\begin{aligned} a_i^* H a_i &= a_i^*(A + I_n)(A - I_n)^{-1} a_i = a_i^*(A + I_n)x = a_i^*(a_i + 2x) \\ &= \|a_i\|_2^2 + 2a_i^* x = 1 - 1 = 0. \end{aligned}$$

Also solved by Herbert Carus and the proposer.

Half Gauss Sums

S 27 [1980, 218]. *Proposed by Emma Lehmer, Berkeley, California.*

Let p be an odd prime and (n/p) be the Legendre symbol. Set $\xi = \exp(2\pi i/p)$. Let Σ denote the sum taken over the range $1 \leqslant n \leqslant (p-1)/2$.

It is well known that the Gauss half-sum $\Sigma(n/p)\xi^n$ has the value

$$\frac{1}{2} p^{1/2} + i\Sigma(n/p)\sin(2\pi n/p) \quad \text{if } p \equiv 1 \pmod 4$$

and has the value

$$\Sigma(n/p)\cos(2\pi n/p) + \frac{1}{2} i p^{1/2} \quad \text{if } p \equiv 3 \pmod 4.$$

Show that

$$\begin{aligned} &(2/p)\Sigma(n/p)\sin(2\pi n/p) < 0 \quad \text{if } p \equiv 1 \pmod 4, \\ &(2/p)\Sigma(n/p)\cos(2\pi n/p) > 0 \quad \text{if } p \equiv 3 \pmod 4. \end{aligned}$$

This problem is solved in an article by Bruce C. Berndt and Ronald J. Evans, *Half Gauss Sums*, Math. Ann., 249 (1980) 115–125. See Theorem 12.

A Fibonacci Sequence of Nested Triangles

S 29 [1980, 302]. *Proposed by Clark Kimberling, University of Evansville.*

Suppose $T = ABC$ is a triangle having sides $AB < AC < BC$ and a point B' on segment BC satisfying $AB' = AB$. Call T *admissible* if the shortest side of triangle $T' = AB'C$ does not touch the shortest side of T, i.e., the shortest side of T' is $B'C$.

(a). Characterize all T for which the sequence $T_1 = T,\ T_2 = T_1',\ T_3 = T_2',\ldots$ consists exclusively of admissible triangles.

(b). For such T, let s_n be the length of the shortest side of T_n and determine $\lim_{n\to\infty}(s_n/s_{n+1})$.

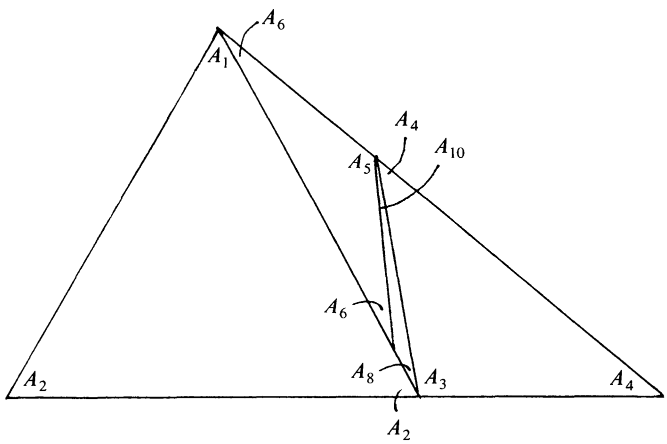
*(c). For such T, let P be the limit point of the nested triangles T_n and determine the angle APB.

Solution to parts (a) *and* (b) *by Clark Kimberling, University of Evansville, Jordi Dou* (*Barcelona, Spain*), *and L. Kuipers* (*Mollens Vs, Switzerland*).

(a) Necessary and sufficient condition is $B/C = (1 + \sqrt{5})/2$ (the golden ratio).

(b) This limit is also the golden ratio.

Proof. (a) Relabel and mark T as follows:



Continuing the notation as indicated, clearly $A_{2n+4} = A_{2n} - A_{2n+2}$ for $n = 1, 2, \ldots$. As a first step in a mathematical induction argument, note that $A_6 = A_2 - A_4$. Assume for arbitrary $n \geqslant 3$ that

$$A_{2n} = (-1)^{n+1}(f_{n-2}A_2 - f_{n-1}A_4),$$

where f_k is the kth Fibonacci number, given by

$$f_1 = 1, \qquad f_2 = 1, \qquad f_k = f_{k-1} + f_{k-2}$$

for $k = 3, 4, \ldots$. Then

$$\begin{aligned} A_{2n+2} = A_{2n-2} - A_{2n} &= (-1)^n(f_{n-3}A_2 - f_{n-2}A_4) - (-1)^{n+1}(f_{n-2}A_2 - f_{n-1}A_4) \\ &= (-1)^n[(f_{n-2} + f_{n-3})A_2 - (f_{n-1} + f_{n-2})A_4]. \end{aligned}$$

Thus $A_{2n+2} = (-1)^{n+2}(f_{n-1}B - f_nC)$ for $n = 2, 3, \ldots$. In the nest of triangles, we are requiring each shortest side to be not touching the shortest side of the preceding triangle. This is equivalent

to $A_2 > A_4 > A_6 > \cdots$, so that (*) for any odd n we must have

$$f_nB - f_{n+1}C > -f_{n+1}B + f_{n+2}C > f_{n+2}B - f_{n+3}C.$$

This implies

$$\frac{f_{n+3}}{f_{n+2}} < \frac{B}{C} < \frac{f_{n+4}}{f_{n+3}}.$$

Since $\lim_{n\to\infty}(f_{n+1}/f_n) = (1+\sqrt{5})/2$, as is well known, we conclude $B/C = (1+\sqrt{5})/2$.

(*) together with the hypothesis $AB < AC < BC$, the inequalities $A_2 > A_4 > A_6 > \cdots$ can be used easily to show that each T_n is a scalene triangle, so that T_n has a unique shortest side.

(b) Each of the triangles T_n has the property just proved for T. That is, $A_{2n}/A_{2n+2} = (1+\sqrt{5})/2$ for $n = 1, 2, \ldots$. This recurrence relation yields $A_{2n} = C^{n-1}/B^{n-2}$. Now using the Law of Sines, we find

$$\begin{aligned}
\lim_{n\to\infty} \frac{s_n}{s_{n+1}} &= \lim_{n\to\infty} \frac{\sin A_{2n+2}}{\sin A_{2n+4}} \\
&= \lim_{n\to\infty} \frac{\sin(C^n/B^{n-1})}{\sin(C^{n+1}/B^n)} \\
&= \lim_{n\to\infty} \frac{\sin x^{-n+1}C}{\sin x^{-n}C} \qquad \text{where } x = \frac{1+\sqrt{5}}{2} \\
&= \lim_{y\to 0} \frac{\sin yC}{\sin yC/x} \\
&= \lim_{y\to 0} \frac{C\cos yC}{\frac{C}{x}\cos yC/x} \\
&= x
\end{aligned}$$

as required.

Solution to part (c) *by Jordi Dou and L. Kuipers*. We note that $\overrightarrow{AP} = \Sigma_1^\infty \overrightarrow{A_iA_{i+1}}$. Let α_i be the angle from $\overrightarrow{A_iA_{i+1}}$ to $\overrightarrow{A_1C_1} = \overrightarrow{AC}$. We have

$$\hat{\alpha}_1 = C_2, \qquad \hat{\alpha}_2 = C_2 + C_3 + \pi, \qquad \hat{\alpha}_3 = C_2 + C_3 + C_4, \qquad \hat{\alpha}_i = \Sigma_2^{i+1}C_i + \pi^{(i)},$$

where $\pi^{(i)} = \pi/2 + (-1)^i\pi/2$. Thus $\hat{\alpha}_i = B(1-\rho^{-1}) + \pi^{(i)}$. The length of A_iA_{i+1} is $|A_iA_{i+1}| = a_{i+2}$, the greatest side of T_{i+2} (or the side b_{i+1} of T_{i+1}). We have

$$\begin{aligned}
a_2/a &= \sin B/\sin A, & a_3/a_2 &= \sin C_1/\sin B, \\
a_4/a_3 &= \sin C_2/\sin C_1, & a_{i+2}/a_{i+1} &= \sin C_i/\sin C_{i-1}, \\
a_{i+2}/a &= \sin C_i/\sin A.
\end{aligned}$$

Thus $a_{i+2} = (a/\sin A)\sin \rho^{-1}B$. Let P', P'' be the projections of P on AC and on the normal to AC from A. Then

$$|AP'| = \left((-1)^{i-1}a/\sin A\right)\sum_1^\infty \sin\rho^{-1}B\cos(1-\rho^{-1})B,$$

$$|AP''| = \left((-1)^{i+1}a/\sin A\right)\sum_1^\infty \sin\rho^{-1}B\sin(1-\rho^{-1})B.$$

The angle $\sphericalangle APB$ can be calculated by using the coordinates of P: (AP', AP''), of B: $(c\cos A, c\sin A)$, and of A: $(0,0)$.

Also solved by Kenneth L. Bernstein, F. S. Cater, and Nick Franceschine III.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by January 31, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2932 **Corrected** [1982, 212]. *Proposed by Henry E. Fettis, Mountain View, CA.*

For $n > 1$, set

$$S_n(b) = \sum_{n=0}^{[(n+b)/2]} (-1)^k \frac{(n-2k+b)^{n-1}}{K!(n-k)!}.$$

Prove that, if $b < n$, $S_n(b) = S_n(-b)$.

E 2956. *Proposed by Barry Powell, Kirkland, Washington.*

Prove that if p is a prime, $p \equiv 1 \bmod 4$, there exists a prime $q < \sqrt{p}$ for which $q^{p-1} \not\equiv 1 \bmod p^2$. (See E 2435 [1973, 943].)

E 2957. *Proposed by Stephen McAdam, The University of Texas at Austin.*

Let f be a function defined on the positive integers. Let $f_1 = f$ and inductively define $f_{k+1}(n) = \Sigma_{d|n} f_k(d)$ over positive divisors d of n. This operation is well known in elementary number theory. Show that if $k > 1$ and if $f_k = f_1$, then f is identically zero.

E 2958. *Proposed by M. S. Klamkin, University of Alberta.*

Let x, y, z be positive, and let A, B, C be angles of a triangle. Prove that $x^2 + y^2 + z^2 \geqslant 2yz\sin(A - \pi/6) + 2zx\sin(B - \pi/6) + 2xy\sin(C - \pi/6)$.

E 2959. *Proposed by Jack Garfunkel, Flushing, New York.*

Triangle ABC is inscribed in a circle. The medians of the triangle intersect at G and are extended to the circle to points D, E and F. Prove: $AG + BG + CG \leqslant GD + GE + GF$.

E 2960. *Proposed by Malcolm J. Sherman, State University of New York at Albany.*

Evaluate

$$\int_0^\infty \frac{dx}{(1+x^2)(1+x^\alpha)}, \text{ where } \alpha \geqslant 0.$$

E 2961. *Proposed by Ivan Niven, University of Oregon.*

Find all triples of positive constants a, b, c such that

$$\lim_{n\to\infty} \left\{ \frac{\sqrt[n]{a} + b}{c} \right\}^n$$

exists (as a finite limit), and determine its value.

SOLUTIONS OF ELEMENTARY PROBLEMS

$3^2 + 4^2 = 5^2$, $3^3 + 4^3 + 5^3 = 6^3$, ...

E 2870 [1981, 148]. *Proposed by Gideon Schwarz, Hebrew University, Jerusalem.*

Find all positive integers n such that $3^n + 4^n + \cdots + (n+2)^n = (n+3)^n$ (Example: $3^2 + 4^2 = 5^2$).

Solution by Hugh M. Edgar, San Jose State University; Lenny Jones, Shippensburg State College; David C. Kurtz, University of Malawi; and E. Trost, Zürich. The given inequality holds for $n = 2, 3$. For $n = 4$ we have $L_n = 3^n + \cdots + (n+2)^n < (n+3)^n$. For $n > 4$, using induction, $L_{n+1} < (n+3)L_n + (n+3)^{n+1} < 2(n+3)^{n+1}$. But $2(n+3)^{n+1} < (n+4)^{n+1}$ when $n = 5$ by calculation, and for $n \geqslant 6$ by an easy manipulation of $[(n+4)/(n+3)]^{n+1} - 2$.

Also solved by U. Abel, M. Ascher, M. D. Ašić (England), K. L. Bernstein, D. M. Bloom, R. Breusch, K. A. Brown, Jr., P. S. Bruckman, T. Bu (Norway), Chico Problem Group, S. Curran, P. N. R. de Souza (student, Brasil), C. W. Dodge, D. Dunham, E. A. Escalona F. (Venezuela), G. Fisher, L. L. Foster, B. J. Gaitley, G. A. Gonnet (Canada), R. P. Grimaldi, G. A. Heuer, D. T. Hùng, E. Johnston, H. Kappus (Germany), G. S. Lessells (Ireland), K. Y. Li (student), J. Lindgren, S. Locke, B. Richmond & B. Richter, O. P. Lossers (The Netherlands), H. M. Marston, R. Moller, D. Moore, W. A. Newcomb, T. D. Ngo, D. A. Rawsthorne, K. Rogers, St. Olaf Problem Solving Group, I. A. Sakmar (Canada), B. M. Scott, I. J. Schoenberg, G. Schwarz (Israel), R. E. Shafer, D. B. Shapiro, G. Shulman, A. Stenger, B. M. Stewart, R. A. Struble, University of South Alabama Problem Group, M. Wolterman, and the proposer.

A Type of Repeating Decimal

E 2871 [1981, 148]. *Proposed by W. G. Leavitt, University of Nebraska.*

Let n be a positive integer relatively prime to 10. Call n a nines number if, for every h relatively prime to n, the decimal expansion of h/n has even period, and the sum of the two half-periods is $999\ldots$. Show that n is a nines number if and only if $10^k \equiv -1 \pmod{n}$ for some k.

Solution by E. Trost, Zürich, Switzerland. Since $(n, 10) = (n, h) = 1$, h/n has a pure recurring decimal expansion δ with a period of just r digits, where r is the order of 10 (mod n).

(a) We assume that there is a k with $10^k \equiv -1 \pmod{n}$, k being the smallest exponent with this property. Then the order of 10 (mod n) is $2k$. Therefore δ has even period and the half-periods A, B consist of k digits. Letting $h < n$ we have $\delta = .ABAB\ldots$ and $10^k\delta = A.BAB\ldots$. Taking into account that $(10^k + 1)(h/n)$ is an integer, we infer $A.BAB\ldots + .ABAB\ldots = A + 1$. This implies $A + B = 99\ldots9$ (k digits).

(b) Now let $\delta = .ABAB\ldots$ and $A + B = 99\ldots9 = 10^k - 1$. Multiplication by 10^{2k} gives $10^{2k}\delta = A \cdot 10^k + B + \delta$. From this follows $\delta(10^{2k} - 1) = A(10^k - 1) + 10^k - 1$ or $(h/n)(10^k + 1) = A + 1$. Therefore $10^k \equiv -1 \pmod{n}$.

Also solved by U. Abel (West Germany), K. L. Bernstein, J. A. Brandler, K. A. Brown, Jr., P. S. Bruckman, B. Cheng & D. T. Hung, Chico Problem Group, L. Erlebach, L. L. Foster, V. Hernandez (Spain), J. P. Hoyt, E. Johnston, K. Y. Li, H. M. Marston, A. Smuckler (Israel), L. Somer, University of South Alabama Problem Group, M. Woltermann, and the proposer.

Foster mentioned that, if $t \geqslant 1$ and if 10 has even order mod p, then p^t is a nines number. Several solvers noted that the product of two nines numbers need not be a nines number.

$\Sigma a_i = \Sigma b_i,\ \Pi a_i = \Pi b_i$

E 2872 [1981, 148]. *Proposed by J. G. Mauldon, Amherst College.*

Find five different triples of positive integers with the same sum and the same product.

Solution by Lorraine L. Foster and Gabriel Robins (Student), California State University at Northridge. The following *ten* triples have sum 1326000 and product $2^7 3^6 5^4 7^2 13^3 17^3$: (83300, 495720, 746980), (79968, 573750, 672282), (80325, 560235, 685440), (143325, 224640, 958035), (139230, 232050, 954720), (119340, 278460, 928200), (106080, 324870, 895050), (92820, 397800, 835380), (89505, 424320, 812175), (79560, 596700, 649740).

These were generated as follows. Using a modest computer and various memory sparing techniques, numbers $s \geqslant 20$ were partitioned into three parts and all products were computed. For $s = 100$, three triples, (6, 45, 49), (7, 30, 63) and (9, 21, 70) were found to have common product $P = 13230$. The equations $2^4 a + b + c = 200$, $bc = P/2a$, were investigated. Two solutions

$(a, b, c) = (1, 49, 135), (9, 21, 35)$ were found. Thus (multiplying our previous triples by 2) we obtained five triples with sum 200 and product 2^3P. Similarly we generated six triples with sum 400 and product 2^6P. Next, considering the equations $13^3a + b + c = 5200$, $bc = 2^6P/a$ we generated seven triples. Continuing in this fashion we obtained the above ten triples.

Comments: It clearly follows that for $n \geqslant 3$ there exist at least ten n-tuples with common sum and common product. Also for $n \geqslant 6$ there are at least fifty-five such n-tuples!

Further computer work determined that 185 is the minimal number which has five partitions into triples with common product. They are (11, 84, 90), (12, 63, 110), (15, 44, 126), (18, 35, 132), (22, 28, 135). Also 400 is the minimal number with six such partitions.

Moreover, our data would seem to indicate that there are infinitely many primitive sets of five (or perhaps any $n \geqslant 5$) triples with common sum and common product.

Any *pair* of integers is uniquely determined by its sum S and its product P.

Also solved by R. Breusch, P. L. Chabot, B. Gaitley, G. S. Lessells (Ireland), O. P. Lossers (Netherlands), D. Moore, I. Rosenholtz, and the proposer.

Property of Central Factorial Numbers

E 2873 [1981, 208]. *Proposed by K. Satyanarayana, Hyderabad, India.*

Let m and n be positive integers, and

$$P_{n,m} = \sum_{k=0}^{[n/2]} (-1)^k \binom{n+1}{k} (n+1-2k)^m,$$

where $[n/2]$ denotes the greatest integer less than or equal to $n/2$. Show that

(i) $P_{n,m} = 0$ for $n = m + 1 + 2p$, $p = 0, 1, \ldots.$
(ii) $P_{n,m} > 0$ for $n = m - 1 - 2p$, $p = 0, 1, \ldots, [(m-1)/2]$.
(iii)* $P_{2n+1,2m+1} = (-1)^{n-m}\binom{2n}{n} f(n)/(2n-1)(2n-3)\cdots(2n+1-2m)$, where $f(n)$ is a polynomial (of degree m) in n with positive coefficients.
(iv)* $P_{2n,2m}$ is given by the same formula.
(v)* From (iii), (iv) it would follow that $P_{n,m} > 0$ when $m < n$, and $(-1)^{(n-m)/2}P_{n,m} > 0$ when $n \geqslant m$, provided $n + m$ is even. Show this directly.

In the above, n, m are positive integers, and $f(n)$ is a polynomial (of degree m) in n with positive coefficients.

Comment by Ira Gessel, Massachusetts Institute of Technology. The central factorial numbers $T(n, k)$ defined by

$$T(n,k) = \frac{1}{k!}\sum_j \binom{k}{j}(-1)^j\left(\frac{1}{2}k - j\right)^n$$

are discussed in J. Riordan's *Combinatorial Identities* (Wiley, New York, 1980), pp. 213–217 and 233–236. It is easy to see that the numbers $P_{n,m}$ are given by

$$P_{n,m} = 2^m(n+1)!T(m, n+1).$$

Properties (i) and (ii) are immediate from the generating function

$$\sum_{m=0}^{\infty} P_{n,m}\frac{x^m}{m!} = (2\sinh x)^{n+1}.$$

Property (iii) is apparently stated incorrectly, as it contradicts (i).

Also solved by L. Kuipers (Switzerland).

$m/n = 1/x + 1/y$

E 2875 [1981, 208]. *Proposed by David Singmaster, Polytechnic of South Bank, England.*

(a) Which rational numbers m/n can be expressed as $1/x + 1/y$ where x and y are positive integers?

(b)* One can define a density on the rationals m/n with m, n positive by considering m/n as an ordered pair (m, n): For any set A of such ordered pairs,

$$\delta(A) = \lim_{x\to\infty} |A \cap \{1,2,\ldots,x\}^2|/x^2,$$

when the limit exists. Is the set of rationals (m, n), with no common factor, of the form $1/x + 1/y$, of density 0?

Solution by Daniel A. Rawsthorne, Wheaton, Maryland. (a) The number m/n can be expressed as $1/x + 1/y$ if and only if there exist divisors d_1, d_2 of n such that $m \mid d_1 + d_2$. ($\Rightarrow$) Assume $m/n = 1/x + 1/y$. Set $x' = x/(x, y)$, $y' = y/(x, y)$. Then $m(x, y)/n = (x' + y')/(x'y')$, with $(x' + y', x'y') = 1$. Thus there is an integer c such that

$$m(x, y) = c(x' + y') = cx' + cy'; \quad n = cx'y'.$$

($\Rightarrow$) is proved. ($\Leftarrow$) Assume $d_1 \mid n$, $d_2 \mid n$, $m \mid (d_1 + d_2)$. Define d_3, k by

$$km = d_1 + d_2, \quad n = d_3 d_1 d_2/(d_1, d_2).$$

Then

$$m/n = 1/x + 1/y, \quad x = kd_3d_1/(d_1, d_2), \quad y = kd_3d_2/(d_1, d_2). \square$$

(b) The proof depends on the result $\tau(n) = \theta(n^\varepsilon)$ for any $\varepsilon > 0$, where $\tau(n)$ is the number of divisors of n. Clearly

$$\delta(A) = \lim x^{-2} \sum_{n\leqslant x} \operatorname{card}\{m/n \mid (m, n) = 1, m/n \text{ is of the form } 1/x + 1/y\}.$$

By part (a), $m/n \in A$ with $(m, n) = 1$ if and only if m, n are given by

$$n = kd_1d_2, \quad m = d_1 + d_2, \quad (d_1, d_2) = (k, d_1 + d_2) = 1.$$

The cardinality is therefore at most $\sum_{k|n} \tau(k)$. Therefore

$$\delta(A) \leqslant \lim x^{-2} \sum_{n\leqslant x} \sum_{k|n} \tau(k) = \lim x^{-2} \sum [x/n]\tau(n) = 0.$$

Also solved by J. A. Brandler, T. E. Elsner, L. E. Mattics, R. W. K. Odoni (U. K.), V. Pambuccian (student, Rumania), Problemlösegruppe, Mannheim (Germany), and M. F. Wyneken.

The Functional Equation $F(z + 1) - F(z) = (z + 1)^{-1}$

E 2876 [1981, 209]. *Proposed by Doug Hensley, Texas A & M University.*

Let $f(n) = \sum_{k=1}^{n} 1/k$, $n \geqslant 1$.

(a) Is there a continuation of f to an analytic function defined on the interval $x \geqslant 0$ such that $f(x + 1) - f(x) = (x + 1)^{-1}$?

(b)* Is there an analytic continuation of f to the complex half-plane $\mathrm{Re}(z) > 0$ such that $f(z + 1) - f(z) = (z + 1)^{-1}$?

Solution by Tom M. Apostol, California Institute of Technology; Miroslav D. Asic, London School of Economics; Bruce C. Berndt, University of Illinois; Paul Chauveheid, University of Liège, Belgium; Chico Problem Group, California; Robert Corless, student, University of Waterloo, Canada; Larry Eifler, UMKC, Kansas City, MO; Michael B. Gregory, University of North Dakota;

Richard Johnsonbaugh, Chicago State University; Kenneth A. Klinger, CNA Insurance, Chicago, IL; O. P. Lossers, Eindhoven Institute of Technology, Netherlands; L. E. Mattics, University of South Alabama; R. W. R. Odoni, University, Exeter, UK; Problemlösegruppe, Mannheim, Germany; and Robert E. Shafer, Palo Alto, CA. The answer to both questions is "yes." The (standard) notation appears in Abramowitz-Stegun, *Handbook of Mathematical Functions*, NBS, AMS no. 55. Set

$$F(z) = \Gamma'(z+1)/\Gamma(z+1) + \gamma = \psi(z+1) + \gamma,$$

where Γ is the gamma function and γ is Euler's constant. Then (i) $F(z)$ is analytic in $\operatorname{Re} z > -1$, and (ii) $F(z+1) - F(z) = (z+1)^{-1}$; in fact $F(z) = \Sigma_{k=1}^{\infty}[k^{-1} - (z+k)^{-1}]$; see the reference.

The most general solution is $G(e^{2\pi i z}) + \Gamma'(z+1)/\Gamma(z+1)$, where G is an arbitrary entire function.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by January 31, 1983. The solver's full post-office address should be on each sheet.

6393. *Proposed by G. A. Edgar, Ohio State University.*

I. Let $D = \{z \in \mathbb{C}: |z| < 1\}$ and $\bar{D} = \{z \in \mathbb{C}: |z| \leq 1\}$. Prove or disprove: If $f_1: \bar{D} \to \mathbb{R}$ is continuous on $\bar{D}$ and superharmonic on D; if $f_2: \bar{D} \to \mathbb{R}$ is continuous on $\bar{D}$ and subharmonic on D; and if $f_1(z) \leq f_2(z)$ for all $z \in D$, then there exists a function $h: \bar{D} \to \mathbb{R}$, continuous on $\bar{D}$ and harmonic on D, with $f_1(z) \leq h(z) \leq f_2(z)$ for all $z \in D$.

II. Prove or disprove: If μ is a positive measure of total mass 1 defined on the Borel sets of $\mathbb{C}$, with $\mu(\{z: |z| > 1\}) = 0$, such that μ satisfies $\int h(z)\, d\mu(z) = h(0)$ for all harmonic functions h: $\mathbb{C} \to \mathbb{R}$, then μ also satisfies $\int s(z)\, d\mu(z) \geq s(0)$ for all subharmonic functions s: $\mathbb{C} \to \mathbb{R}$.

6394. *Proposed by Jan Mycielski and Andrzej Ehrenfeucht, University of Colorado, Boulder.*

Let n and $m \geq n+1$ be given. Let $X \subseteq R^n$ with $|X| = m$ be a set in general position, i.e., if $Y \subseteq X$ and $|Y| \leq n+1$, then Y spans a $(|Y|-1)$-dimensional hyperplane and no two such hyperplanes are parallel to each other provided that neither of the corresponding two Y's is a subset of the other. Let $f_n(m)$ be the number of linear orderings of X which can be obtained by a perpendicular projection of X into any directed line which is in general position relative to X.

(a) Prove that $f_2(m) = 2\binom{m}{2}$ and $f_3(m) = 6\binom{m}{4} + 4\binom{m}{3} + 2$.

(b) Is $f_n(m)$ well defined (i.e., the same for all X in general position) for $n \geq 4$, and, if so, can it be evaluated?

6395. *Proposed by P. M. Cohn, Bedford College, London, England.*

Let R be a commutative ring and $A = (a_{ij})$ an $n \times n$ matrix over R. Let S be the ring generated by R and n^2 indeterminates x_{ij} such that, on writing $X = (x_{ij})$, the relations $AX = XA = I$ (in matrix form) hold.

Is S necessarily commutative?

6396. *Proposed by N. A. Marlow, American Telephone and Telegraph Company, and M. Tortorella, Bell laboratories.*

Suppose F is a right-continuous probability distribution function on $\mathbb{R}$ with $\mu = \int_{\mathbb{R}} x\, dF(x)$, $-\infty < \mu < \infty$. Put

$$F_1 = F, F_2(x) = \int_{\mathbb{R}} F(x-u)\, dF(u) = F * F, \text{ and } F_{n+1} = F_n * F, n = 2, 3, \ldots.$$

a. If $\mu \neq 0$, show that $\sum_{n=1}^{\infty}[F_n(n\mu) - F_{n+1}(n\mu)]$ (1) diverges.
b. If $\mu = 0$ and $\sigma^2 = \int_{\mathbb{R}} x^2 dF(x) < \infty$, show that (1) converges.

6397. *Proposed by Edgar Feldman and Michael Vulis, CUNY Graduate Center, New York.*

(i) Let $D^n = \{x \in R^n \mid \|x\| < 1\}$ be the unit disk, and let f: $D \to D$ be a diffeomorphism of D such that
(a) $f(f(x)) = x$ for every $x \in D$, i.e., $f \circ f = \text{id}$;
(b) there exists a neighborhood U of 0 in D such that $f|_U = \text{id}$.
Show that $f = \text{id}$.
(ii)* Is the theorem true when f is not a diffeomorphism, but a homeomorphism?

6398. *Proposed by Edgar Feldman and Michael Vulis, CUNY Graduate Center, New York.*

Is it true that for any group G of cardinality less than or equal to that of the continuum, one can construct a metric space M such that G is its full group of isometries? *What if $|G|$ has greater power?

SOLUTIONS OF ADVANCED PROBLEMS

Nesting Regular *n*-gons

6062 [1975, 1016; 1977, 578; 1981, 152]. *Proposed by B. H. Voorhees, University of Alberta.*

Consider an infinite sequence of regular n-gons such that each $(n+1)$-gon is contained within the preceding n-gon and is of maximal area consistent with this constraint. Take the first element of this sequence as an equilateral triangle having unit area. Is the limit of this sequence a point or a circle? If it is a circle, determine its area.

D. J. Daley, Australian National University, has pointed out an oversight in the paper of B. A. Troesch, which contained a proposed solution of this problem (J. Optimization Theory and Applications, 31 (1980, pages 7–16). In D. J. Daley, *Optimally nested regular n-gons*, J. Optimization Theory and Applications (paper accepted for publication) it has now been proved that, as he had conjectured previously, the ratio of the radius of the limit circle to that of the circumscribing circle of the initial equilateral triangle equals 0.1764671 (cf. (0.341473)/2 = 0.170737 as claimed by Troesch), and hence the area sought by Voorhees equals 0.0753105.... Daley's partial solution assumed that each pair of successive regular n-gons $\{P_n\}$ should have a common axis of symmetry. This fact has now been established, by algebraic methods, and $\{P_n\}$ can be so constructed that P_{2n}, P_{2n+1} and P_{2n+2} have a common axis of symmetry, but not P_{2n-1}, P_{2n} and P_{2n+1}. P_{2n} has a pair of opposite sides each with both end-points touching sides of P_{2n-1}, while both end-points of each of two sides almost diametrically opposite of P_{2n+1} touch P_{2n} in two pairs of opposite sides of P_{2n}.

Best Rank-*k* Approximation for a Matrix

6125 [1976, 818; 1980, 495]. *Proposed by Simeon Reich, Tel Aviv University, Tel Aviv, Israel.*

For a given $n \times n$ complex matrix A of rank r, and an integer $k, 1 \leq k \leq r$, a best rank-k approximation of A is a matrix $A_{(k)}$ of rank k satisfying $\|A - A_{(k)}\| = \inf\{\|A - X\|$: X is an $n \times n$ matrix of rank $k\}$ where $\|A\| = (\text{trace } A^*A)^{1/2}$.

Show that if A is normal, then $A^j_{(k)}$ is a best rank-k approximation of A^j for all $j \geqslant 1$, but that this is no longer true for arbitrary A. (Cf. Ben-Israel and Greville, *Generalized Inverses*, Wiley, New York, 1974, p. 250.)

Counterexample by A. J. Bosch, Technological University, Eindhoven, The Netherlands. It is not even true for A normal. Counterexample: Take $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. A is symmetric, hence normal. $A_{(1)} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ is a best rank-1 approximation, but $A^2_{(1)} = 0$ so cannot be a best rank-1 approximation for A^2.

The same idea gives counterexamples for all values of n, and for $0 < k \leqslant n - 1$.

A Nonlinear Iteration

6304 [1980, 582]. *Proposed by I. J. Schoenberg, University of Wisconsin.*

If $0 < t_1 < t_2 < 1$, $r > 0$, $s \geqslant 1$, set

$$t_1^1 = t_1 + t_1^r(t_2 - t_1)^s$$
$$t_2^1 = t_1 + t_2^r(t_2 - t_1)^s$$

and iterate to obtain

$$t_1^{(n+1)} = t_1^{(n)} + \left[t_1^{(n)}\right]^r\left(t_2^{(n)} - t_1^{(n)}\right)^s,$$
$$t_2^{(n+1)} = t_1^{(n)} + \left[t_2^{(n)}\right]^r\left(t_2^{(n)} - t_1^{(n)}\right)^s,$$

$n = 0, 1, \ldots$. Show that $\lim t_1^{(n)} = \lim t_2^{(n)}$ and if possible evaluate the common limit.

Solution by Paul S. Bruckman, Concord, California. We first note that there is a slight error in the statement of the problem. The left members of the first two equations should be $t_1^{(1)}$ and $t_2^{(1)}$, not t_1^1 and t_2^1. We make a change in notation for convenience. Let

$$x_n = t_2^{(n)},\ y_n = t_1^{(n)},\ (n = 0, 1, 2, \ldots),\ \left(\text{where } t_1 = t_1^{(0)},\ t_2 = t_2^{(0)}\right). \tag{1}$$

Thus, the given recursions become:

$$x_{n+1} = y_n + x_n^r(x_n - y_n)^s, \tag{2}$$

$$y_{n+1} = y_n + y_n^r(x_n - y_n)^s, \qquad n = 0, 1, 2, \ldots. \tag{3}$$

We first prove the following inequalities:

$$0 < y_0 < y_1 < y_2 < \cdots < y_n < x_n < x_{n-1} < \cdots < x_2 < x_1 < x_0 < 1. \tag{4}$$

Let S denote the set of nonnegative integers n such that (4) holds. By hypothesis, $0 \in S$. Suppose $k \in S$. Using the induction hypothesis, (2) and (3), we have:

$$x_{k+1} - y_{k+1} = (x_k^r - y_k^r)(x_k - y_k)^s > 0.$$

Thus $x_k > y_k \Rightarrow x_{k+1} > y_{k+1}$.

Again using the induction hypothesis and (2), along with the given condition $s \geqslant 1$, $x_{k+1} < y_k + 1^r(x_k - y_k) = x_k < 1$; also, using (3), $y_{k+1} > y_k > 0$.

It follows that $k \in S \Rightarrow (k + 1) \in S$. By induction, (4) is established.

Now (x_n) and (y_n) are monotonic bounded sequences, hence have limits, say x and y, respectively. Furthermore, we see from (4) that

$$0 < y < 1. \tag{5}$$

Taking limits in (3), we then have

$$y = y + y^r(x - y)^s \Rightarrow y^r(x - y)^s = 0 \Rightarrow x = y. \tag{6}$$

In the general case a closed form expression for x does not seem to be readily available. If $r = 1$, however, we may express x as an infinite product. Define

$$d_n = x_n - y_n,\, d = d_0 = x_0 - y_0 = t_2 - t_1. \tag{7}$$

Then, subtracting (3) from (2), we obtain:

$$d_{n+1} = d_n^{s+1}. \tag{8}$$

It follows by an easy induction that

$$d_n = d^{(s+1)^n}. \tag{9}$$

We note that

$$y_{n+1}/y_n = 1 + d_n^s. \tag{10}$$

Forming products on both sides of (10) by letting n approach infinity, we obtain the limiting expression (which we know exists):

$$y = y_0 \prod_{n=0}^{\infty} \{1 + d^{s(s+1)^n}\}. \tag{11}$$

In the particular case $r = s = 1$, the expression in (11) may be evaluated in closed form, for then

$$y = y_0 \prod_{n=0}^{\infty} (1 + d^{2^n}) = y_0 \prod_{n=0}^{\infty} \left\{\frac{1 - d^{2^{n+1}}}{1 - d^{2^n}}\right\} = y_0(1 - d)^{-1}, \quad \text{or}$$

$$y = y_0(1 - x_0 + y_0)^{-1}. \tag{12}$$

Also solved by L. Kuipers (Switzerland) and Michael Skalsky.

A Set of Homogeneous Equations

6312* [1980, 675]. *Proposed by M. S. Klamkin, University of Alberta.*

Prove or disprove that the set of n equations in n unknowns

$$x_1^{l_i} + x_2^{l_i} + \cdots + x_n^{l_i} = 0 \qquad (i = 1, 2, \ldots, n),$$

where the l_i's are relatively prime positive integers, has only the trivial solution $x_i = 0$ $(i = 1, 2, \ldots, n)$, if and only if each $m = 2, 3, \ldots, n$ divides at least one l_i.

Solution by Constantine Nakassis, Gaithersburg, Maryland. Let $n > 2$ be an even number $(n = 2k)$; suppose that the only even number in $l_1, l_2, \ldots, l_n$ is l_1 (for example, take $l_1 = n!$ and let $l_2, \ldots, l_n$ be the first $(n - 1)$ primes that follow n). Consider any k complex numbers which satisfy

$$y_1^{l_1} + y_2^{l_1} + \cdots + y_k^{l_1} = 0.$$

Let $x_{2i-1} = y_i, x_{2i} = -y_i$ for $i = 1, 2, \ldots k$. It is clear then that the proposed system has nontrivial solutions. (The starred assertion is true if $n = 2$, but false if $n = 2k + 1 > 3$.)

The case $n = 3$ remains open; the starred assertion has been established by the proposer for many triples.

ANSWERS TO "PHOTOS" ON PAGE 455

Top: Israil Moiseevich Gelfand; bottom: Mark Aronovich Naimark.

Multiplication Table of S_n

6318 [1980, 758]. *Proposed by Cole A. Giller, University of California, Berkeley.*

Let M_n be the $n! \times n!$ matrix obtained from the multiplication table of the full symmetric group S_n by replacing each entry s of the table by 1 if s is an n-cycle, and by 0 otherwise. When is $\det(M_n) \neq 0$?

Solution by D. J. Kleitman, Massachusetts Institute of Technology.

1. The matrix in question is (by definition) upon relabelling the rows by the inverse of the corresponding entry, the matrix of the sum of the n-cycles in the regular representation of S_n.
2. The sum of the n cycles commutes with all group elements and is diagonal if one chooses a basis such that the regular representation is reduced to irreducible ones.
3. The diagonal elements of each is (by definition) a nonzero multiple of the character of the given irreducible representation evaluated on the n-cycle class. The determinant is nonzero if and only if all these characters are nonzero.
4. This holds for $n \leqslant 3$, but no larger n. The characters of the symmetric group can be calculated easily and explicitly by the Murnaghan-Nakayama rules (see Littlewood, *Theory of Group Characters*) and characters found that vanish on n-cycles for any n, $n \geqslant 4$. An easy way to see this is to recall that the characters are integers and must obey the orthoquality relation

$$\sum_{\chi} \chi(g)\bar{\chi}(g) = |S_n|/|C| = n = \sum_{\chi} \chi^2(g)$$

for $g \in C$, C the class of n-cycles. Thus at most n characters can fail to vanish on this class and there are more than n irreducible representatives for $n > 3$.

This latter argument was submitted by the proposer Cole Giller, who attributes it to David Richman. C. Greene supplied the former.

Integral Curves of Differential Equations

6321 [1980, 759]. *Proposed by N. P. Erugin, Minsk, USSR.*

Give a qualitative description of the integral curves of the system of differential equations

$$\dot{x} = y - x + x^3, \quad \dot{y} = -x - y + y^3.$$

Find a periodic solution.

Partial solution by Gilbert N. Lewis, Michigan Technological University. First, there is only one critical point $(0,0)$ of the system. This is also a periodic solution. This fact can easily be seen since the graphs of $y = x - x^3$ and $x = -y + y^3$ intersect only at the origin.

Let $r^2 = x^2 + y^2$ and $\tan\theta = y/x$. Then

$$2r\dot{r} = 2x\dot{x} + 2y\dot{y} = 2\big(x^2(x^2 - 1) + y^2(y^2 - 1)\big).$$

This is negative if $r < 1$. Also, if $r^2 = x^2 + y^2 > 2$, then $x^2 > 2 - y^2$ and $x^2 - 1 > 1 - y^2$, so that

$$2r\dot{r} > 2\big((2 - y^2)(1 - y^2) + y^2(y^2 - 1)\big) = 4(1 - y^2)^2 \geqslant 0.$$

Thus, if $r > \sqrt{2}$, then $\dot{r} > 0$.

We also see that

$$\dot{\theta} = \big(-(x^2 + y^2) + xy(y^2 - x^2)\big)/(x^2 + y^2) = -1 - r^2 \sin 4\theta/4.$$

This is negative for $r < 2$, so that θ decreases for $r < 2$. Thus, the solution curves spiral in toward the origin inside the circle $r = 1$. Also, $\dot{\theta}$ is negative except inside and on the curves $r^2 \sin 4\theta = -4$, which are hyperbola-like curves in each quadrant, which become asymptotic to the lines $\theta = \pi/4$

$+ n\pi/2$ and $\theta = n\pi/2$, $n = 0, 1, 2, 3$. Therefore, the solution curves will spiral outward, clockwise, until they enter one of the regions described above, inside which the solution curves approach the line $\theta = n\pi/2$ ($n = 0, 1, 2$ or 3), with θ increasing toward $n\pi/2$ and r increasing without bound.

This analysis also shows that there is at least one unstable limit cycle which lies between the circles $r = 1$ and $r = \sqrt{2}$, and any nonclosed solution curve lying close to a limit cycle will spiral away from it in a clockwise direction.

Also solved by R. A. Struble.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Plane Algebraic Curves: An introduction via valuations. By Grace Orzech and Morris Orzech. Marcel Dekker, New York, 1981.

DOUGLAS COSTA
Department of Mathematics, University of Virginia, Charlottesville, VA 22903

Suppose you were going to write a book on algebraic geometry. How would you do it? Where would you begin? How much or how little of the subject would you attempt to include? What background would you assume on the part of the reader? At what level of sophistication would you write? Introductory? Graduate student? Young researcher? Jaded expert? And even more important, what language would you use? Would you use a classical geometric language with a little algebra thrown in? Would you use the classical algebraic language of function fields, with no explicitly geometric objects in view? Would you use the classical mixed language of affine and projective varieties and their coordinate rings? Or would you choose a more modern sheaf-and-scheme-theoretic approach, or perhaps the more extreme approach of representable functors? Or would you write a book on several complex variables?

Clearly all these decisions, and more, would have to be made before you began writing. It is enough to make most people turn from the task, and perhaps explains why most prefaces do not begin "I wrote this book because I felt the urge to write an algebraic geometry text," but more often begin, "This book grew out of a course given. . . ." In other words, these books tend to get written after the critical decisions have been forced upon their writers by having had to present the material to a given group of people at a given moment in their mathematical lives. And indeed the book under review here, *Plane Algebraic Curves* by Orzech and Orzech, is an example of this type. It is based on a one-term course designed to follow an introductory course on algebraic number theory, including some of the theory of valuations and general Dedekind domains. In fact, the book is subtitled "An introduction via valuations," and valuations are effectively used throughout the book to reduce many of the proofs to arithmetical computations. This is one of the very appealing features of the book. Computations are at the heart of mathematics, and it behooves a writer on algebraic curves to do enough of them to give the reader a feeling for exactly how they enter into the subject. To be sure, many of the computations in a particular subject are routine, dull, and boring. On the other hand, there are those instances in which one senses that if only he knew how to do *this* calculation, great insight would follow. The Orzechs have managed to write an account in which those key computations are explicitly carried out, while most of the "routine" ones are left to the reader. Writing in this way eliminates the clutter of detail which can

distract a reader from the main questions, and allows for plenty of descriptive prose in which to sketch the broad outlines of a subject, showing where it is heading, and how it will get there. This adds enormously to the readability of a book, and is all to the good. Despite several apologies for "vagueness," the Orzechs have done this, and the result has been an attractive introduction to algebraic curves.

Now one may take issue with a judgment of readability for a book which begins with 16 pages of prerequisites, but there are prerequisites and prerequisites. (Compare with the 126 pages of "foundational material" in Griffiths and Harris's epic work [**3**].) The assumption of some familiarity with Dedekind domains, valuations, and the like, is reasonable for most mathematicians, at least those who call themselves algebraists, and for many graduate students who have had a basic algebra course. And, too, the language of modern algebra and valuation theory systematizes and streamlines much of the theory, making it seem more natural and less ad hoc than it might in an elementary account "from scratch." So the prerequisites in this case have the happy effect of rendering the presentation clearer, while preserving its accessibility to a fairly wide audience. (The extent to which it is accessible to undergraduate students is something I will address later, in a more general context.)

Thus, the Orzechs' book provides a case in point, an example of how an author's answers to our initial questions affect the comprehensibility, readability, and ultimate value of his work. The problem of comprehensibility and readability in mathematical writing is, unhappily, a recurring one, and has been reemphasized in some recent commentary (see for example [**6**], [**7**]). It seems ironic that this should be so when one considers that the chief purpose of mathematics is communication. Mathematics is essentially an extension of language designed to enable people to communicate complex ideas in a humanly comprehensible form. Its aim is to constantly develop new symbols (or new uses for old symbols), and rules for their use, which will give us the ability to express compactly ideas whose expression in ordinary language might not be within the limits of human capacities, or even possible at all. And mathematics does this fantastically well. Think for a moment of how long it would take to write out entirely in English prose a complete description of the meaning of, say Euler's formula $e^{i\pi}+1=0$, or the Cauchy integral formula $f(z)=(1/2\pi i)\int_C(f(\zeta)/(\zeta-z))d\zeta$. One can call up a multitude of examples to illustrate this point, but one that has a particular immediacy for me is Green's theorem:

$$\oint_{\partial D} P dx + Q dy = \int\int_D \left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right) dx\, dy.$$

Each time I finish teaching the standard one-semester several variable calculus course which ends with this theorem, I am astonished to realize that most of the term has been devoted to explaining the meaning of the symbols in this single equation.

But what of this view of mathematics as an extension of language, a form of communication? What about all the powerful mathematicians at work transforming famous conjectures into celebrated theorems? Isn't that what mathematics is? Superficially, the answer is yes; big problems and big solutions are the current events of mathematics. But in a deeper way I believe the answer must be no. Consider, for example, the Newton-Leibniz calculus. What has been its principal contribution to civilization? It has not been the solution of any one particular problem of celestial mechanics, of electro-magnetism, of economics, or what have you. Its significance lies in the fact that it empowered mankind to solve a wide variety of problems by providing a language for the analysis of change. The solution of the brachistochrone problem by the Bernoullis was a major event of its day, but in the longer perspective of history we see its significance as marking the birth of the calculus of variations. Likewise in recent times the solution of the four-color conjecture by Appel and Haken was a dramatic achievement, but I venture to guess that the lasting importance of the problem will be in the advances in discrete mathematics which the search for its solution engendered.

So big problems are distinct from mathematics itself, but play a central role in the development

of new mathematics. They do this in large part by stimulating the intellects of mathematicians. Mathematicians are the keepers of mathematics, those who nurture it, push it to its utmost limits, make it grow, and keep it alive to be passed on to new generations. And big problems and big solutions are the fuel that fires their engines. A good problem can launch mathematicians on a crusade, and big solutions are those intermittent victories without which it would be impossible to continue the quest, constant reminders of the effectiveness of mathematics and the power of mathematicians. The dynamic interactions between these three elements, mathematics, mathematicians, and problems, are what drive the development of mathematics.

But I have digressed from the main question of comprehensibility and readability in mathematical writing. Because it is the job of mathematicians to create new ideas and new forms of expression with which to analyze them, and because even the creators of these things do not often fully understand them, writing comprehensibly about them for a larger audience is a built-in difficulty. But it is just that, a difficulty, not an impossibility. What are some of the reasons mathematicians have failed to overcome it? Surely there are those cases in which elitism, and the attendant egotism, are at work. These are the instances in which a flurry of unnecessary symbolism, neologisms, and citations to impenetrable works, serve to put on display the depth and breadth of the writers's knowledge, at the expense of letting his reader in on what he is trying to say. I believe that this kind of elitism is relatively rare, however, as a cause of bad writing. More common, perhaps, is simple laziness. Mathematicians working in a particular field commonly have a working language in which they talk and think. It is simply easier for them to use this language again when they write, whether it is appropriate or not. When this is done deliberately, it is laziness. When it is done unconsciously, it is what is sometimes called "tunnel vision," innocent enough, but no less damaging to the readability of the writing produced. I believe this "tunnel vision," or self-absorption, to be the most common cause of poor mathematical writing. An example of this sort of absent-mindedness occurs in *Plane Algebraic Curves*. In a text otherwise devoid of the language of categories, the correspondence between varieties and their coordinate rings is presented as a fully faithful covariant functor, with no explanation of the meaning of these terms. This has a jarring effect, even upon a reader familiar with such things. I must point out, however, that this small aberration is the only such lapse in this generally well-written book. The lesson in all this is that good writing requires painstaking care. It deserves as much effort and thought as the mathematics it is intended to convey.

The last issue I wish to discuss here is the place of such subjects as plane algebraic curves in the undergraduate curriculum. There are two main questions here: Is it accessible to undergraduates, and should it be taught? The Orzech's book is, I think, not quite accessible to undergraduate students (exceptional students excluded, of course). Even after a standard course in modern algebra, most undergraduates would not have enough background for the presentation via valuations, or the sophistication to match the style in which the book is written. On the other hand, a zealous and vigorous instructor blessed with enthusiastic pupils could make it accessible by beginning with a crash course in valuation theory. But this begs the general question. I believe that the subject of plane algebraic curves is by its very nature accessible to undergraduate students. What makes it possible to offer undergraduate courses in elementary differential geometry is the fact that all mathematics majors have a background in calculus, upon which such courses are traditionally based. What makes it possible to offer elementary number theory courses is the fact that almost all people have a background in arithmetic. And likewise I believe that it should be possible to offer courses in plane algebraic curves because of the fact that most students are already familiar with and have strong intuitive feeling for them. After all, a plane algebraic curve is nothing more than a graph, and everyone who has been through high school has drawn plenty of them. Say "$y = x^2$" to someone, even a mathematician, and odds are that the first image that pops into his head will be the graph of that equation. This indicates the presence of beliefs and intuitions strong enough to sustain a course. And perhaps it indicates also that a good approach to the subject would be to invert the usual order of things. Instead of teaching a course

in plane curves based on algebra, in which the algebraic objects are imbued with the greater reality, regarded as more familiar, and looked upon as sources of information about curves, perhaps it would be better to teach a course based on curves and use them as an invitation to algebra. (Cf. [**4**], especially the introduction.)

On the question of whether plane algebraic curves should be taught to undergraduates, I think it would be mistaken to argue that among the topics of algebraic curves, number theory, or non-Euclidean geometries, for example, one was more intrinsically important than another. What I do feel strongly, however, is that such "pure" or "cultural" subjects should continue to be offered to mathematics majors, and perhaps even be required of them. There is currently a strong movement toward "applied" mathematics and away from "pure" mathematics. In many ways this may be seen as a healthy sign, but in undergraduates who lack a liberal education in mathematics, this perspective can become distorted. My personal experience with mathematics majors today is that many of them have come to identify "applied" mathematics with marketable skills. To an alarming degree many are interested in learning statistics, numerical analysis, and computer languages to the exclusion of such things as complex variables, measure theory, group theory and the like. These latter things they regard as too pure, too abstract, and not applicable. How they have gotten this idea I do not know, but something should be done to dispel it.

Good education sometimes calls for a bit more authoritarianism on the part of faculties than is currently in fashion. Let us not forget that undergraduates come to the university in ignorance. They do not know what it is that they need to know, and it is up to the faculties to tell them. The pressures which have led away from this are continuing. Dwindling numbers of students, increasing numbers of faculty, eroding financial support for higher education, and increasing student preoccupation with careers are pushing universities ever closer to becoming centers for vocational training, or perhaps more accurately for vocational accreditation. Faculties, more and more eager to please in hard times, are evolving into "service providers," their students into consumers of education. To these consumers an old warning is appropriate: caveat emptor.

References

1. G. Orzech and M. Orzech, Plane Algebraic Curves, Marcel Dekker, New York, 1981.

2. W. Fulton, Algebraic Curves, W. A. Benjamin, New York, 1969.

3. P. Griffiths and J. Harris, Principles of Algebraic Geometry, Wiley, New York, 1978.

4. A. Seidenberg, Elements of the Theory of Algebraic Curves, Addison-Wesley, Reading, MA, 1968.

5. R. J. Walker, Algebraic Curves, Princeton University Press, Princeton, NJ, 1950.

6. R. P. Boas, Can we make mathematics intelligible?, this MONTHLY, 88 (1981) 727–731.

7. H. B. Griffiths, Review of A Formal Background to Mathematics by R. E. Edwards, Bull. Amer. Math. Soc., 5 (1981) 354–358.

A Simple Non-Euclidean Geometry and its Physical Basis. By I. M. Yaglom. Translated from the Russian by Abe Shenitzer, with the editorial assistance of Basil Gordon. Springer-Verlag, New York, 1979. xviii + 307 pp.

E. G. STRAUS
Department of Mathematics, University of California, Los Angeles, CA 90024

F. Klein's Erlanger Programm was a fundamental attempt to put some systematic order into the variety of geometries that had been developed in the 19th century. The significance of the program for geometry goes much deeper. Every geometric concept has an associated group of transformations which preserve it, and it gives rise to a more general concept under any larger group of transformations. Thus the Erlanger Programm is a guide to deeper insights and new discoveries.

As an illustration, consider the obvious theorem of Euclidean geometry: *The locus of the midpoints of parallel chords of a circle is the diameter perpendicular to the chords.* Here "midpoints" and "parallel" are invariants under general affine transformations, while "circle" and "perpendicular" are invariants under similarity transformations. It therefore is natural to look at the affine content of the theorem. Here "circle" generalizes to "ellipse" and "diameter perpendicular to the chords" to "diameter joining points at which the tangent is parallel to the chords." Thus the Euclidean theorem generalizes to the affine theorem: *The locus of the midpoints of parallel chords of an ellipse is the diameter of the ellipse which joins the points at which the tangents of the ellipse are parallel to the chords.* If we escalate the group of transformations from affine to projective, the "parallel chords" become "chords on a pencil of concurrent lines"; "midpoints" become "fourth harmonics" (with respect to the endpoints of the chord and the point of the pencil); "ellipse" becomes "conic section"; and the "diameter which joins $\cdots$" becomes "chord which joins points of the conic section at which the tangents are elements of the pencil." Thus the Euclidean theorem gains the projective setting: *The locus of fourth harmonics on a family of chords of a conic section which lie on a pencil with point* 0, *is the chord of the conic section which joins the points where the tangents pass through* 0.

This progress from the special to the general concept and from the special to the general theorem in mathematics, plays an even more fundamental role in all of theoretical science: "It is a wonderful feeling to perceive the unity of a complex of phenomena, which appear as entirely separate things to the observations of the senses." (Einstein—at 22—in a letter to Marcel Grossmann.) This is the way Einstein stated the central goal of scientific theory. He incidentally points up one of the great difficulties, that the "perceived unity" may appear to go counter to the "observation of the senses" (mathematical intuition, etc.). The "correct" concepts and laws come from the demand for unity, not merely by an accumulation of observations.

Invariance in Euclidean geometry is the study of rigid bodies, as far as the congruence theory is concerned; and the study of faithful maps as far as the similarity theory goes. The group of motions is implicit rather than explicit, though Yaglom makes it clear how this group gives a preferred role to the main curves of Euclidean geometry, the straight line and the circle, as the curves which have "glides," that is, a transitive family of rigid motions, onto themselves. That is the way he (no doubt correctly) interprets Euclid's: *A straight line is a line which lies evenly with the points on itself.*

Galileo's principle of relativity (inertia) is a beautiful blend of experience and theoretical insight as stated in his Dialogue Concerning The Two Chief World Systems: "Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies, and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drops fall into the vessel beneath; and, in throwing something to your friend, you need throw it no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. In jumping, you will pass on the floor the same spaces as before, nor will you make larger jumps toward the stern than toward the prow even though the ship is moving quite rapidly, despite the fact that during the time that you are in the air the floor under you will be going in a direction opposite to your jump. In throwing something to your companion, you will need no more force to get it to him whether he is in the direction of the bow or the stern, with yourself situated opposite. The droplets will fall as before into the vessel beneath without

dropping toward the stern, although while the drops are in the air the ship runs many spans. The fish in their water will swim toward the front of their bowl with no more effort than toward the back, and will go with equal ease to bait placed anywhere around the edges of the bowl. Finally the butterflies and flies will continue their flights indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship, from which they will have been separated during long intervals by keeping themselves in the air. And if smoke is made by burning some incense, it will be seen going up in the form of a little cloud, remaining still and moving no more toward one side than the other."

Contrast this with the more theoretical "thought experiment" and the more explicit search for a "guiding principle" that Einstein describes as leading him to his principle of (special) relativity (Autobiographical Notes 1949): "The longer and more desperately I tried, the more I came to the conviction that only the discovery of a general formal principle can lead to assured results. As a model I saw thermodynamics. The general principle there was given by the postulate: The laws of nature are such that it is impossible to construct a *perpetuum mobile* (of the first or second kind). How is such a principle to be found? Such a principle arose from a paradox which I encountered already at the age of 16: When I pursue a light ray with speed c (speed of light in a vacuum), then such a light ray would be perceived by me as a resting, spatially oscillating, electromagnetic field. Such an object does not appear to exist either on the basis of experience or according to Maxwell's equations. It seemed intuitively clear to me in advance, that to such an observer everything must follow the same laws as to an observer at rest relative to the ground. For how should the first observer know, or be able to ascertain, that he was in a state of rapid uniform motion?" Galileo distills his principle of "no absolute uniform motion" from a variety of mechanical observations. Einstein postulates his principle as a guide in the search for correct concepts and laws, willing to abandon such "intuitively obvious" concepts as "distance" and "simultaneity" in the process.

In the study of non-Euclidean geometries one does not pass from the group of Euclidean isometries to groups containing the Euclidean group, but rather to different groups of transformations. However, one is still guided by the analogy with the (implicit) uses of the group structure in Euclidean geometry. Just as the natural curves there are the orbits under one-dimensional subgroups, the straight line and the circle, so in every geometry these orbits play a basic role and in many ways behave analogously to the straight lines and circles of Euclidean geometry. In the geometries studied by Yaglom the groups are affine or projective so that the straight lines are invariants and the innovations relate to the "circles" of the new geometries.

The main content of Yaglom's book is the geometry of the Galilean plane with one time dimension x and one space dimension y invariant under the uniform motions

$$x' = x + a$$
$$y' = vx + y + b.$$

The development is primarily guided by the analogies with the concepts of Euclidean geometry, but the reader is reminded periodically of the mechanical meaning of the geometry in terms of motions on a straight line. Even with the author's masterful exposition, one gets a feeling of nostalgic envy at the thought that there exists a high school class which is able to follow these lectures.

The Lorentz transformations of special relativity and the Minkowskian plane are developed in a separate chapter, with an illuminating section on the Galilean plane as a limiting case of the Euclidean and Minkowskian planes.

The book concludes with three appendices describing the nine geometries—here called Caley-Klein geometries—obtained by combining the three metrics, elliptic, parabolic and hyperbolic on the (real or projective) line with the three dual metrics on the pencil of lines. The analytic description is given in two ways. First, by axiomatizing three-dimensional real vector spaces and then defining the invariant quadratic forms and the points and lines of the various planes. Second, by introducing algebras

$$x + iy, x + \varepsilon y, x + ey, i^2 = -1, \varepsilon^2 = 0, e^2 = 1$$

on the plane and describing the geometries in terms of these algebras. While the exposition is elementary, a good deal of knowledge would be needed before any but the most exceptional reader can appreciate the content of these sections. This masterfully lucid book would make an excellent starting point for an undergraduate seminar in geometry.

Discovering Relativity Theory for Yourself. By Sam Lilley. Cambridge University Press, New York, 1981. $49.50 (hard cover); $19.95 (paperback).

VADIM KOMKOV
Department of Mathematics, West Virginia University, Morgantown, WV 26506

If a book appears entitled *Relativity for the masses*, or *Discovering relativity*, the obvious question should be asked: Do we need another popular exposition? Many excellent popular texts exist: G. Gamow's *Mr. Tomkins in Wonderland* written for school children, G. B. Rumer's *What is relativity*?, Albert Einstein's *Relativity*: *The special and general theory* are all written for nonexperts who are not familiar with differential geometry. At a higher level of sophistication, Hermann Weyl's *Was is materie*?, V. L. Ginsburg's *About the theory of relativity* (in Russian), and A. Einstein's *The meaning of relativity* provide excellent introductory texts.

The present book is different. It is almost a programmed instruction manual organized into 33 chapters, each chapter broken into short sections. Lilley argues very effectively against the ingrained prejudice of the average man. It is hard to comprehend spacetime, and accept that "Space tells matter how to move, and the matter tells space how to curve." The deeper one gets into implications of spacetime geometry, the less unintuitive become the predictions. The existence of black holes follows as a direct consequence of assigning an appropriate metric (distance) to spacetime. Gravitational waves are predicted by the theory (without much experimental data to confirm their existence) and the universe is classified into possible types which may exist.

The usual explanations for the necessity of introducing new ideas into physics at the beginning of the 20th century generally start with the Michelson-Morley experiment which suggested that the speed of light is invariant. That is, an observer travelling toward the source of light or away from it will still measure the same velocity of light C. If one accepts this experimental result, it becomes necessary to question the metaphysical concepts of absolute space and absolute time. Newton's laws, using the terms "a body is at rest, or it travels in a continuous straight line motion with constant velocity," imply the existence of some absolute system of reference. Even there, we must stipulate that such straight line motion in a coordinate system X' will not appear to be curved in our system X. That is, only *inertial* systems are considered. Clearly, any coordinates attached to the earth's surface will not constitute such inertial system. (Recall the Foucault pendulum experiment!) The fact that C appears to be constant to any inertial observer is used to explain slowly and in elementary terms that as two observers measure the same events by means of light signals, each will question the clock (time measurements) and distances measured by the other one. Apparent paradoxes, like the twin paradox, when one twin ages much faster by staying at home, the difficulty of deciding which of two events occurs first, arise naturally from the implication that C is not additive ($C + V \stackrel{?}{=} C$). These paradoxes are beautifully illustrated in Gamow's book.

It is hard to believe at first that the speed of light is the same, irrespective of the motion of the observer. However, the Michelson-Morley and other experiments confirm this fact, and our beliefs have to accept physical reality.

Consider observers A and B who conclude that at a certain time their distance to event P is

equal to their distance to event Q. They may observe a supernova flash in distant parts of the universe. However A and B have relative velocity v. A is stationary relative to P and Q, while B travels toward the supernova Q. According to A the events P and Q were simultaneous. According to B, Q is farther away than P, since he is travelling toward Q, therefore he was nearer to P when the explosion occurred, while his own speed does not add to the speed of the light signals from either P or Q. Hence, the idea of "simultaneous events" depends on the observer. Moreover, as the observers measure distances and times elapsed between different events, they will come to entirely different conclusions. Let us take the speed of light to be a unit of velocity. We consider observers A and B recording some events O and Q. For the sake of simplicity let B be present at O and Q when they occur. According to B he is stationary. It is A who appears to be dashing around the universe at speed v. B concludes that the events O and Q happened at the same place (since he was there and did not go anywhere); hence the spacial distance ΔX between O and Q is zero. According to the observer A, the observer B is confused. While A is standing still, B is travelling at speed v, say, towards Q. It is obvious to A that the events O and Q occurred in different locations. In fact he knows how to measure the distance between them by observing on his clock how long it took B to get from O to Q. To those who have had some modern physics the answer is well known—the time and space measurements of A and B are related to each other by the Lorentz transformation. For those who have had none, I recommend the careful introduction given by Lilley to the spacetime measurements for observers travelling with constant speed relative to each other, or accelerating relative to each other. His discussion of the Special Theory of Relativity, when you have inertial observers in a gravity-free space, is slow, careful, and follows the Socratean principle of asking questions rather than answering them.

Most nonexperts still find Einstein's unification of space and time into spacetime puzzling. Einstein immediately realized that spacetime must have local curvature. This is a difficult concept to absorb. If one looks at a smooth curve in the x-t coordinates, various properties of the curve depend on our choice of an orientation of the coordinate system. The rate of change dx/dt clearly depends on such choice. But the curvature $d\theta/ds = \rho^{-1}$ (θ is the angle between tangent direction to the curve and an arbitrary fixed direction, ds is infinitesimal increase of length along the curve) only depends on our choice of scale (units of s) but not, for example, on the orientation of the coordinate axes. Here one quantity ρ^{-1} defines the curvature. In spacetime we need 20 quantities at each point to define the curvature (the components of the Riemann-Christoffel tensor). Reducing this number to 10 (components of the Ricci tensor) Einstein related the properties of this tensor to the rates of change of energy in spacetime. The details are impossible to explain in simple terms.

But let us return to Lilley's book. The cover review is right, but only partially. The first part of the book—the first six chapters—can be read by a reader without any mathematical preparation. After that the uninitiated person would have to be very smart to follow the theory. Occasionally, even the expert mathematician may be confused if he does not read carefully every word. There is an unnecessary amount of elementary explanation. Why explain that $3 - 5 = -2$ to a reader who is later asked to comprehend the significance of the covariant nature of Ricci's tensor, or to argue (using commutativity of multiplication of real numbers) that $(ab)^2 = a^2b^2$, or that multiplication is distributive over addition, and then expect the same reader to understand manipulations involving infinitesimal increments. It amounts to an assumption that the reader is a very ignorant genius. For me it is hard to imagine that a reader capable of comprehending the implications of the zero geodesic property along signal lines would not know how to add negative and positive numbers. The author does not pontificate but argues with the reader, issuing numerous challenges. "Prove that for yourself!" "What else do you see?", "Ponder on the last sentence," "Do you find it puzzling?" are examples of bold print challenges to the reader at the end of a section. The chapters 1 to 6 contain only descriptive material. The author even explains that 5^2 means 5 times 5, and also explains Pythagoras' theorem about right triangles.

The book is written with a tremendous enthusiasm by a man who loves the subject and is good

at explaining it. I would not hesitate to recommend the first ten chapters to any group of readers with no mathematical background, who wish to acquire an understanding of the concept of the special theory of relativity. It should provide deep enjoyment, intellectual stimulation, and admiration for the man who is responsible for this beautiful theory.

The Mathematical Experience. By Philip J. Davis and Reuben Hersh. Boston: Birkhauser, 1981. xix + 440 pp. $24.00.

WILLIAM M. PRIESTLEY
Department of Mathematics, University of the South, Sewanee, TN 37375

"Statesmen despise publicists, painters despise art critics, and physiologists, physicists, or mathematicians have usually similar feelings; there is no scorn more profound, or on the whole more justifiable, than that of the men who make for the men who explain. Exposition, criticism, appreciation, is work for second-rate minds." So wrote G. H. Hardy at the beginning of *A Mathematician's Apology*, one of the most beautiful books in or out of mathematics, and a book in which a first-rate mind turned his attention to exposition, criticism, and appreciation of mathematics. Since that time, fortunately, other first-rate minds have been drawn in the same direction, exposing many different views of the nature of mathematics. The latest effort along these lines, by Philip J. Davis and Reuben Hersh—though not offering a view that Hardy would admire—is perhaps the most ambitious and best of all in describing what mathematicians really do. Lest this essay, in focussing as it must on a small part of *The Mathematical Experience*, mislead, let me state plainly at the outset that every student of mathematics will find this book valuable, and that Davis and Hersh, who write as one, doubly deserve our gratitude.

What is mathematics, and what manner of spirit animates its study? That, of course, is a philosophical question. One of many favors done us by Davis and Hersh is to clear the air a bit by pointing out that the so-called philosophies of logicism, formalism, and intuitionism are not and never have been philosophies of mathematics. Their focus, of course, has always been on the foundations of mathematics, and they became known as "philosophies" only because there was a period in the early part of this century when almost all work in the philosophy of mathematics was work in foundations. Despite its great importance, the study of foundations is only one of a multitude of interesting aspects of mathematics.

In fact, the diversity of interesting aspects is bewildering, as we find ourselves saying that mathematics is like X, like Y, and like Z, when X, Y, and Z themselves are quite dissimilar. Davis and Hersh give us a number of arguments, each well done, from the totality of which we must conclude that mathematics is like an ideology, a religion, or an art form, and is thus a humanistic study, "one of the humanities"; and yet mathematics has a science-like quality in that "its conclusions are compelling, like the conclusions of natural science."

Is it inconsistent to say that one entity, mathematics, can resemble each of so diverse a collection of bedfellows? Perhaps not. Davis and Hersh remind us that sometimes the only way to know a complicated object (for example, mathematics) is to describe it by making (necessarily incomplete) analogies with objects that are already familiar. It is, after all, a standard procedure to study planar sections of a complicated geometric object, or to study representations of a complicated algebraic object. The inconsistent appearance of different sections or representations is more to be expected than not. Therefore, when we attempt to describe mathematics by making analogies, we should not be surprised to come up with analogies that appear to be inconsistent. It is possible that each of many incomplete analogies may be realized, provided we do not lock ourselves into a rigid point of view. It is possible that the logicists, formalists, intuitionists,

Platonists, constructivists, etc., are all talking about the same thing when they speak of mathematics. Perhaps there is a valid conception of mathematics that is large enough to be, as mathematics sometimes seems to be, all things to all men.

The conception of mathematics offered by Davis and Hersh is contained in the following statement.

> [T]he study of mental objects with reproducible properties is called mathematics.

This concise statement is an appealing definition of mathematics, although it can hardly reconcile all camps. The adjective "reproducible" allows mathematics to be like a science, and the adjective "mental" forces mathematics to be a human activity—one of the humanities—with the possibility of continually enriching human culture. By a "mental object" is meant, for example, an idea, but Davis and Hersh apparently wish to avoid the word "idea" because of its Platonic connotations of eternal existence independent of human consciousness. While they are not willing to go that far, they do take a step in Plato's direction by insisting that our interests lie in properties reproducible (by others), and therefore mathematical objects are indeed independent of *individual* human consciousness. Davis and Hersh conclude that mathematical objects exist in the collective consciousness of (the mathematically literate segment of) society. While the conclusion is surely not new (one suspects that sociologists have for years catalogued mathematics as belonging to something like the *conscience collective* of Emile Durkheim) it is interesting that we here have a definition that implies this conclusion.

"Mathematics, being a human activity," they say, "...profits greatly from individual genius, but thrives only with the tacit approval of the wider community. As a great art form, it is humanistic; it is scientific-technological in its applications." Mathematics is thus caught in the struggle between the individual and society as well as the struggle between the arts and sciences. In the arts-science tension there is nothing which would surprise us. Mathematics draws vitality from being stretched on one side toward beauty, form, and vision; on the other toward utility, function, and rationality.

The surprise comes when Davis and Hersh discuss how the individual mathematician contributes to society's growth of knowledge. Isn't it done by giving, for example, a full and infallible proof of a significant new assertion? Not at all, say Davis and Hersh, following Imre Lakatos in condemning the rigid formalist approach to mathematics "which tends to identify mathematics with its formal axiomatic abstraction." Indeed, one cannot point to a full proof of any significant mathematical assertion, if "full" demands the exclusive use of symbols of first-order predicate logic. Thus, virtually all real proofs of mathematics are informal to some degree, and Davis and Hersh argue that if they are informal to any degree, then they are fallible. To abandon so easily the notion of absolute rigor in "real" mathematics is disquieting, to say the least, but it is true that the use of a nonformal language can lead to unforeseen misunderstandings. I remember an algebra text that began by briskly reviewing properties of the integers and the reals. On the same page it was asserted that every nonempty set of positive integers has a least element, and that every bounded nonempty set of real numbers has a greatest lower bound. In context, the first "has" means "contains"; but the second does not. To miss the distinction is to confuse the discrete with the continuous. I wanted to write the author about this small obfuscation, but a colleague convinced me that if a reader did not understand what the author meant, then the reader was unqualified to open the book in the first place. Davis and Hersh would probably agree.

> The actual situation is this. [In real mathematics] proofs ... are established by "consensus of the qualified." A real proof is not checkable by a machine, or even by any mathematician not privy to the gestalt, the mode of thought of the particular field of mathematics in which the proof is located. Even to the "qualified reader," there are normally differences of opinion as to whether a real proof (i.e., one that is actually spoken or written down) is complete or correct. These doubts are resolved by communication and explanation, never by transcribing the proof into first-order predicate calculus. Once a proof is "accepted," the results of the proof are regarded as true (with very high probability)....

This passage implies not only that absolute rigor is unattainable in real mathematics, but also that the formulation of a proof and its acceptance by the mathematical community is of necessity a social process. Formalism, which assumes absolute rigor, is therefore not a valid model of what real mathematicians do. (Formalism is still, of course, of great interest in itself, and perhaps might be regarded as a model of what the "ideal" mathematician does.)

Whether one sides with Davis, Hersh, and Lakatos or not, one must admit that the implications of accepting formalism as an honest philosophy of mathematics are hard to take, both in terms of the validity of the formalist description and the values implicit in it. Often a theorem of mathematics can be described in first-order logic no better than a newborn baby can be described in terms of protons and electrons. All significance having been lost, one would never guess that mathematics had something to do with solving problems. Moreover, the axiomatic method casts undue emphasis upon the individual steps of a derivation, thereby deemphasizing the value of general methods of proof and virtually ignoring the construction of examples and counterexamples as a valid mathematical pursuit. How can one know the significance of a theorem without knowing examples of its applicability and without knowing enough about potential counterexamples to appreciate the tack taken by its proof? And with the possible exception of a devotee of parlor games, what human being is interested in insignificant mathematics?

Indeed, the question "What is significant mathematics?" is at least as important as the question "What is mathematics?"—and one may spend some time wondering in which order these questions should be attacked. It may seem foolhardy to attack the former question before the latter (although one can, for example, intelligently discuss "significant figures" without discussing "figures"), but the reverse order may be just as foolhardy. Both questions could be settled at once by a definition of mathematics that has within it a (perhaps necessarily vague) implication of value, to guide us in judging the significance of a proffered piece of mathematical literature. The simultaneous settling of these notorious questions, however, appears to be nearly as far off as the settling of the seemingly parallel questions of aesthetics, viz., "What is art?" and "What is significant art?"

These questions may not, in fact, be parallel, for art and mathematics intersect in several places. The most interesting and most mysterious is in the realm of creativity. Davis and Hersh agree with others about the creation of new mathematics in saying that *it does not come from the mind alone*. Without the body and its senses, many of the elements of creative mathematical thought—visual, tactile, kinesthetic, even muscular, according to Hadamard's well-known survey—would not exist. The mind alone, even Descartes's mind, which tried to separate itself from the body that made it, is too small to encompass the mathematical experience.

Davis and Hersh discuss the nature of the creative act of the individual, but they appear to prefer to emphasize the role played by mathematicians collectively. This emphasis is probably justified in view of the relative lack of attention given, until lately, to the collective or social role. *The Mathematical Experience* comes to a close by leaving us in the cultural heights of the *conscience collective*.

It is exhilarating to think of the spirit of mathematics dwelling here in the high reaches, promoting that dispassionate, rational outlook without which the shared understanding of our discipline would not be possible. And it may well be that future generations will increasingly accept mathematics as being animated by a cultural force acting on a global level. But such a spirit can be acknowledged only in the abstract manner of a philosopher who reifies the "Zeitgeist" of an era during which no one dreamed of a *Geist* in his *Zeit*.

We should not forget that there is in mathematics a spirit that is the direct opposite of an abstraction. It is so real, in fact, that it even cries out and is heard, though only on auspicious occasions. It is the spirit that was never so much at home as when it resided within the body of Archimedes and raised the roofs of Syracuse with its colossal shouts of surprise and delight.

Writing about Archimedes, Plutarch described this spirit in a phrase that has been rendered into English both as a "raging Siren" and as a "familiar demon." The first is an apt description of

the overflow of exuberance in the moment of light; but such moments come only to those who can stand the dark. The second seems more descriptive of that spirit of compelling total engagement whose charm kept Archimedes through the long nights and made him

> forget his food and neglect his person, to that degree that when he was occasionally carried by absolute violence to bathe or have his body anointed, he used to trace geometrical figures in the ashes of the fire, and diagrams in the oil on his body, being in a state of entire preoccupation, and, in the truest sense, divine possession....

One suspects that all mathematicians have known all manner of approximations to this kind of spirit, and that little nontrivial mathematics has been done without its aid. Isn't this, then, what we should mean when we speak of the spirit of mathematics? If so, how does it enter into the general picture of things?

The picture of mathematics given us by Davis and Hersh is of something that begins with an individual and journeys to the collective consciousness, undergoing metamorphoses along the way. Others, including myself (not necessarily in disagreement, but with symmetry as a dominant sense) would prefer to say that the journey also ends with an individual—the reader.

> [W]hat happens to a proof when it is believed? The most immediate process is probably an internalization of the result. That is, the mathematician who reads and believes a proof will attempt to paraphrase it, to put it in his own terms, to fit it into his own personal view of mathematical knowledge. No two mathematicians are likely to internalize a mathematical concept in exactly the same way, so this process leads usually to multiple versions of the same theorem, each reinforcing belief, each adding to the feeling of the mathematical community that the original statement is likely to be true.
>
> —DeMillo, Lipton, and Perlis, Social processes and proofs of theorems and programs, Comm. ACM, 22 (1979) 271–280.

If we see mathematics as going from individuals to individuals, rather than from individuals to society, and if we see creative mathematics as coming from the body (not just from the mind, but from all the senses as well), then there is a suggestion of a physiological—rather than sociological—basis for the mathematical activity of human beings. Such a basis for creative activity in general, and for poetry in particular, has been proposed by Stanley Burnshaw in an intriguing book called *The Seamless Web* (Braziller, 1970). However speculative or implausible a physiological basis for mathematics may at first appear, it might help to explain why mathematics sometimes appears to be not willed by the mind, but quite the opposite—ideas popping up unexpectedly out of nowhere. At such times we are inclined to say that mathematics is something done to a mathematician rather than something done by a mathematician. These disparate feelings might be harmonized by adopting the view that mathematics is a result of the interaction of an organism, a human being, with its environment. In such an interaction it is natural to find that neither side will totally predominate. It is also natural (rather than metaphysically impossible) to find that mathematics bears a close relationship to our environment. And our most vivid mathematical experience has been that, since the time of Pythagoras, problems without have somehow become problems within, burdening us with such stirrings that we sometimes physically feel their presence.

The Greeks, perhaps in response to this violence within, coined the phrase *en theos*, now familiar to us as *enthusiasm*, which means "a god inside." This, it seems to me at least, is about the best we can do when we are asked what is the spirit of mathematics. This is what tells us that it is of the nature of a human being to do mathematics, just as it is of the nature of an organism to strive to assimilate a compelling presence. What else can explain, following assimilation, our primal cry?

Bounded Analytic Functions. By John B. Garnett. Academic Press, New York, 1981. xvi + 467 pp., \$59.00.

PETER W. JONES
Department of Mathematics, University of Chicago, Chicago, IL 60637

During the past 25 years there has been a tremendous amount of activity centered around the study of H^p spaces. What are these spaces and what properties do they have? A function f is in the Hardy space H^p, $0<p<\infty$, if f is analytic on the open unit disk, Δ, of C, and

$$\sup_{0<r<1}\left(\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^p d\theta\right)^{1/p}\equiv\|f\|_{H^p}<\infty.$$

For $p=\infty$, we define H^∞ to be the space of all functions f which are analytic on Δ and satisfy

$$\sup_{z\in\Delta}|f(z)|\equiv\|f\|_{H^\infty}<\infty.$$

By an old theorem due to Fatou, every function $f\in H^\infty$ has radial boundary values almost everywhere on the unit circle, T, in the sense that $f(\theta)=\lim_{r\to 1}f(re^{i\theta})$ exists a.e. $d\theta$. Furthermore, f may be recaptured by taking the harmonic extension of the boundary values $f(\theta)$. It turns out that for any value of p, H^p functions still have radial boundary values almost everywhere, and when $p\geqslant 1$, the function can be recaptured from the boundary values. (When $0<p<1$, the boundary values must be considered not in the almost everywhere sense, but rather in the sense of distributions.) Let us look at three sample problems that illustrate some of the themes of current research on H^p spaces.

When $1\leqslant p\leqslant\infty$, the space H^p is a Banach space. What is the dual of H^p? When $p=\infty$, the dual space is a very large and intractable space which we will not discuss. When $1<p<\infty$ it turns out that the dual space of H^p is isomorphic to H^q, where $(1/p)+(1/q)=1$. What makes life interesting is that the dual of H^1 is not H^∞. Let us see how we can find the dual of H^1. Now the dual space of $L^1\equiv L^1(T)$ is $L^\infty\equiv L^\infty(T)$, and by using boundary values we can show that $H^1\subset L^1(T)$ and that $\|f\|_{H^1}=\|f(\theta)\|_{L^1}$ if $f\in H^1$. By the Hahn-Banach theorem $(H^1)^*$ must be L^∞ modulo the annihilator, i.e., those functions $\phi\in L^\infty$ such that $\int_0^{2\pi}f(\theta)\phi(\theta)\,d\theta=0$ for all $f\in H^1$. It turns out that the annihilator is H_0^∞, the class of all H^∞ functions vanishing at the origin. Thus $(H^1)^*\cong L^\infty/H_0^\infty$.

What makes H^1 even more interesting is the fact that there is another representation for $(H^1)^*$ which seems completely different from L^∞/H_0^∞. Let BMOA (for analytic bounded mean oscillation) be the class of all functions ϕ in $H^1(T)$ which satisfy

$$|\phi(0)|+\sup_I\frac{1}{|I|}\int_I|\phi(\theta)-\phi_I|\,d\theta\equiv\|\phi\|_*<\infty.$$

Here the supremum is taken over all arcs $I\subset T$, and

$$\phi_I=\frac{1}{|I|}\int_I\phi(\theta)\,d\theta$$

is the mean value of ϕ on I. It is a remarkable fact, due to Charles Fefferman, that $(H^1)^*=\text{BMOA}$ under the mapping $f\to(1/2\pi)\int_0^{2\pi}f(\theta)\,\overline{\phi(\theta)}\,d\theta$. This result has several important implications. The first is that BMOA is very often much easier to work with than L^∞/H_0^∞; this leads to a powerful tool to work with when studying H^1. A second pleasant consequence is that difficult results about H^p, when $1<p<\infty$, can be reduced to much easier problems about BMOA. An example of the second phenomenon would be the study of the boundedness of translation invariant operators on H^p spaces.

Now that we have looked at H^1, let us turn to H^∞. What does an H^∞ function look like? Before giving one answer to this problem, let us take a look at some examples. Suppose $\alpha\in\Delta$ and

$f(z) = (z - \alpha)/(1 - \bar{\alpha}z)$. It is not hard to compute that $\| f \|_{H^\infty} = 1$. Suppose we now multiply together functions of the above type. Let $\{\alpha_n\}$ be a sequence of points in $\Delta \setminus \{0\}$ and let

$$B(z) = z^m \prod_{n=1}^{\infty} \left(\frac{z - \alpha_n}{1 - \bar{\alpha}_n z} \cdot \frac{\bar{\alpha}_n}{\alpha_n} \right).$$

Then each factor of B has H^∞ norm 1. One can show that this product converges to a nonzero H^∞ function if and only if $\{\alpha_n\}$ satisfies the Blaschke condition: $\sum_{n=1}^{\infty} (1 - |\alpha_j|) < \infty$. In that case B is called a Blaschke product, and one can show that $|B(\theta)| = 1$ a.e. $d\theta$. While it is easy to find H^∞ functions which are not Blaschke products, one might take a wild guess that every H^∞ function is almost (in some sense) a sum of Blaschke products. This wild guess turns out to be true as Donald Marshall's beautiful theorem asserts: The closed convex hull of the Blaschke products is the unit ball of H^∞. What this means is that if $\| f \|_{H^\infty} < 1$, then there are sequences $\{\lambda_j\}$ of complex numbers and $\{B_j\}$ of Blaschke products such that $\sum_{j=1}^{\infty} |\lambda_j| \leq 1$ and $f = \sum_{j=1}^{\infty} \lambda_j B_j$.

The space H^∞ occupies a special role in the H^p theory because it is not only a Banach space, but it is closed under pointwise multiplication of elements, and $\| fg \|_{H^\infty} \leq \| f \|_{H^\infty} \| g \|_{H^\infty}$. In other words, H^∞ is a (commutative) Banach algebra. For this reason it is interesting to look at the maximal ideal space of H^∞, $\mathfrak{M}$, i.e., the collection of all maximal ideals of H^∞. One can show by functional analysis that every maximal ideal corresponds to a (unique) complex homomorphism of H^∞ and vice versa. For this reason, $\mathfrak{M}$ is often identified with the collection of all complex homomorphisms. What do points in $\mathfrak{M}$ look like? Notice that if $z_0 \in \Delta$ and $\phi: H^\infty \to C$ is defined by $\phi(f) = f(z_0)$, then $\phi(f + g) = \phi(f) + \phi(g)$ and $\phi(fg) = \phi f(f)\phi(g)$, i.e., ϕ is a complex homomorphism. It is not hard to show that there are points in $\mathfrak{M}$ which do not correspond, in the above manner, to points in Δ. Now $\mathfrak{M}$ turns out to have a natural topology on it which is called the Gelfand topology. A natural conjecture is that $\mathfrak{M}$ is the closure of Δ in the Gelfand topology; this is true and is known as Lennart Carleson's corona theorem. Using functional analysis one can show that proving the corona theorem is equivalent to solving the following problem. Suppose $f_1, \ldots, f_N \in H^\infty$ and suppose there is $\varepsilon > 0$ such that $\sum_{j=1}^{N} |f_j(z)| > \varepsilon$ for all $z \in \Delta$. Construct $g_1, \ldots, g_N \in H^\infty$ such that $\sum_{j=1}^{N} f_j(z) g_j(z) \equiv 1$ on Δ. What makes this last problem especially interesting is that is is not a problem in functional analysis, but rather a problem in function theory: how does one construct H^∞ functions?

In order to understand the solutions of the above problems, one must first be more familiar with the structure of H^p functions. The first three chapters (131 pages) of Garnett's book are devoted to such preliminaries and can be understood by anyone familiar with basic results in real and complex analysis and functional analysis. The other seven chapters are devoted to more advanced topics and the book covers almost all of the important advances the subject has made during the past 25 years. There is also much in the book that is new. The book is beautifully and clearly written. Each chapter ends with a long list of problems (many with extensive hints) which cover several results not presented in the text. For the student or researcher who wants to be able to follow current research in H^p theory, this is the perfect book to read. It is a safe bet that the 1980's will see a large number of graduate students "cut their teeth" on Garnett's book.

MISCELLANEA

80. There is a strong movement afoot to increase the amount of secondary material included in a course for the purpose of linking it up with other subjects to which it may apply. So strong has this tendency become that there is some danger of its being overdone. Indeed, accusations have been made that some text-books and some courses consist entirely of so-called motivating material to the exclusion of tangible principles on which to base the work.—E. L. Mickelson, this MONTHLY, 38 (1931) 454. (From a book review.)

TWO-YEAR COLLEGE MATHEMATICS READINGS

Edited by Warren Page

viii + 304 pages. Paperbound
List $19.50   MAA Member: $14.50

This work surveys the two-year college mathematics landscape. Stops are made at points of interest that cover areas from algebra to recreational mathematics.

There are many things of interest for everyone in this book. Some are expository articles that introduce the reader to new advances in mathematics. Other articles can be used to enrich course work and to stimulate both the student's and instructor's mathematical intellect.

This book will serve as an excellent source of enrichment for the talented student. Teachers of undergraduates will find it is an excellent reference source that they will want to consult when preparing lectures.

Topics covered are:

- **Algebra**
- **Geometry**
- **Number Theory**
- **Calculus**
- **Probability and Statistics**
- **Calculators and Computers**
- **Mathematics Education**
- **Recreational Mathematics**

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 8 October 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

EDITOR

Main articles, editorial correspondence, and reprint permission: P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

THE WIDTH OF A CHAIR

GILBERT STRANG

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

Moving a large chair is not much fun. That is especially true when a door is involved. Probably the designer verifies that his chair will go through a standard door, but he never supplies instructions, and trial and error is the algorithm most frequently adopted. At present we do not know a uniformly successful alternative, but at least we can call attention to the problem and give a partial solution. In the case of a convex chair, the problem is easier but still not completely solved.

We begin in two dimensions, with a compact set C and the closed interval $I = \{(x, y): x = 0, 0 \leq y \leq l\}$. The goal is to determine necessary and sufficient conditions for C to pass through I by a continuous family R_t of rigid motions—translations combined with rotations. It is like mailing a postcard of shape C into a slot I, and the motion can bring points of C back through the slot before the whole set ultimately passes through. For a convex set intuition suggests that a reversal of this kind, and even a rotation, should not be necessary: if the set can get through at all, it can do so without turning. We anticipate a single rotation, to make C as thin as possible in the vertical direction, and a single translation to put C to the left of I. Then if C can pass through I, translation in the x direction should do it. Our contribution is to confirm that this intuition is correct in the plane.

In three dimensions it is not correct. Harold Stark has constructed convex chairs which can pass through a door, either square or circular, although no projection of the chair will fit in the doorway. The chair itself can go through, but it cannot be put into a box (or a cylinder) that will. I am extremely grateful to the referee who observed that my original example was unsatisfactory, and to Harold Stark and Mike Artin for correcting it. The smallest box into which it will fit is not known.

1. Convex Sets in R^2. Suppose C is a compact convex set in the plane. The thickness in the direction of any unit vector θ is the distance between the two tangent lines, one on each side of C, which are perpendicular to θ. In terms of the support function $s(\theta)$, this distance is

$$d(\theta) = s(\theta) + s(-\theta) = \max_{u \in C} \langle u, \theta \rangle - \min_{v \in C} \langle v, \theta \rangle.$$

In other words, $d(\theta)$ is the length of the orthogonal projection of C onto a straight line parallel to θ. The minimum of $d(\theta)$ is the *width* w of the convex set C, and it is attained by a "perpendicular chord"; there is a chord of length w which is perpendicular to the boundary at both ends [**1**]. It is this chord which seems likely to become stuck if the interval I is too small, but our proof is more indirect.

THEOREM 1. *A compact convex set C can pass through I if and only if $w \leq l$.*

Proof. If $w \leq l$, then C goes through by pure translation as described above. It only needs to be properly positioned, and then we can find a family R_t that moves it from the left halfplane to the right:

$$R_0C \subset \{x < 0\}, R_1C \subset \{x > 0\}, R_tC \cap \{x = 0\} \subset I \quad \text{for all } t. \tag{1}$$

Gilbert Strang received his Ph.D. in 1959 under the guidance of Peter Henrici at UCLA. He has worked since then at M.I.T. His research led to a book with George Fix on *An Analysis of the Finite Element Method*, and the Association awarded him the 1976 Chauvenet Prize for a paper on that subject. His text on *Linear Algebra and Its Applications* grew out of his teaching at M.I.T. His current interest is in optimal design, including the shapes evolved by biological structures, and the applications of duality to mechanics and to continuous networks. The present paper came from trying to move a chair onto a hotel balcony (it went through but wouldn't go back).

For the converse, assume first that the boundary ∂C is smooth; through every boundary point there is a unique tangent line, and it varies continuously along ∂C. Suppose that C can pass through I (with rotations allowed), and consider at each instant t the chord $R_tC \cap I$; it is the intersection of chair and doorway (or postcard and slot). At the ends of the chord we construct the tangent lines to R_tC. In Fig. 1a the set is just entering the interval, and in 1b it is leaving:

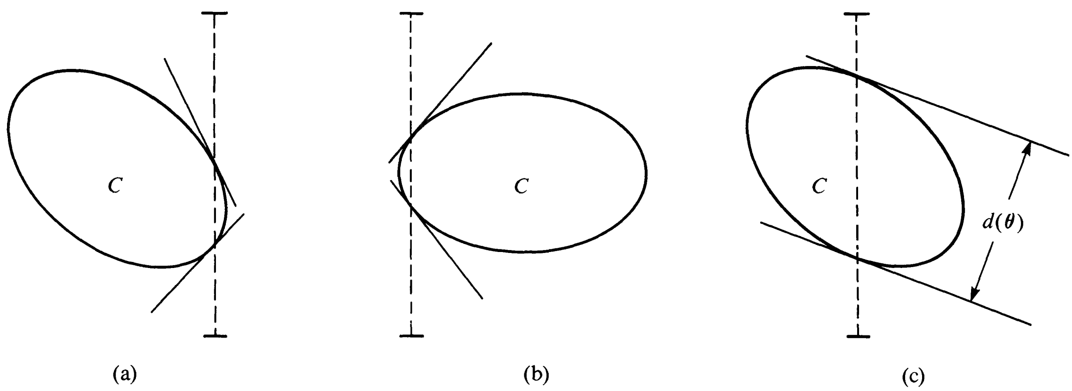


FIG. 1

The slopes of these tangents are continuous functions of t. Therefore there will be an instant at which the two tangents are parallel (Fig. 1c). At that point the chord $R_tC \cap I$ has length not more than l and not less than the distance $d(\theta)$ between the tangent lines. Therefore

$$l \geqslant d(\theta) \geqslant w.$$

The conclusion remains true for a convex set with corners; one possibility is to introduce a sequence of smooth convex subsets C_n converging to C. As C passes through I so do the C_n, and their widths must satisfy $w_n \leqslant l$. Therefore $w \leqslant l$, and the theorem is proved. Another possibility is to accept that the tangent may not be unique at all boundary points, and to go more carefully through the original argument.

2. Convex Sets in R^3. A first step is to establish the condition under which a convex set can pass through a strip of thickness l, say the strip

$$S = I \times (-\infty, \infty) = \{(x, y, z) \colon x = 0, 0 \leqslant y \leqslant l, -\infty < z < \infty\}.$$

The width of C is again the least distance between supporting planes.

THEOREM 2. *C can pass through S if and only if $w \leqslant l$.*

The proof is not changed; we consider the cross section $R_tC \cap S$ cut out by the strip, and construct the two vertical planes that are tangent to R_tC at the cross section. As the set enters the strip, these planes coincide with the y-z plane. By the time the set emerges, each plane has turned through 180°, and they have done so in opposite senses. Therefore there is an instant at which they are parallel, with the set between them, and Figure 1c represents the projection onto the x-y plane. It yields $w \leqslant l$ as before, which is therefore necessary as well as sufficient for C to pass through S.

We intended to reduce the strip to a finite door, and establish that C could pass through only if its projection onto some plane fits into the door. However, that is not true; it is possible that a convex set can get through with rotation, and not without. Our first example fits a pyramid through a circular door (a porthole) by tilting it to let the vertex through. Then we pass a tetrahedron through a square door with two turns.

Stark's pyramid is easy to visualize, but comparatively uncomfortable as a chair. Its base is a square of side 2, equal to the diameter of the door. Its height exceeds 1, but the vertex is

sufficiently off-center that the base can go through first. In Fig. 2a, the set is moved to the right until the trapezoid $ABTS$ reaches the doorway. Then tipping the pyramid allows the vertex to pass.

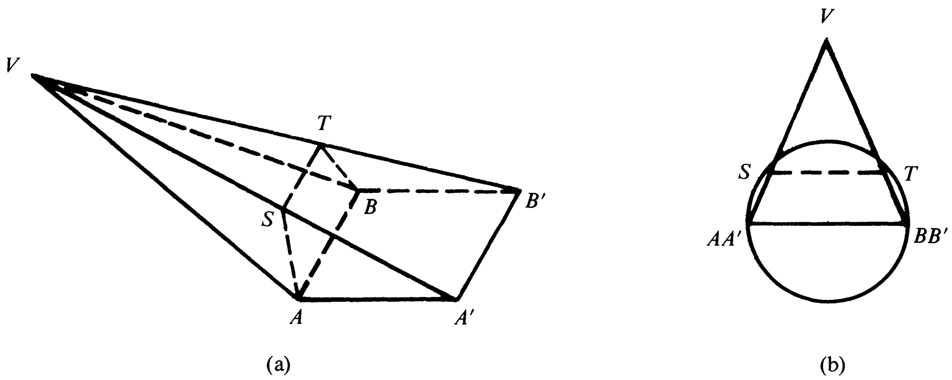


FIG. 2. (a) The turning point $ABTS$. (b) Front view.

We note that no projection can fit into the circular opening. The base itself will barely fit, and only when it is projected onto a diameter. (The diagonals of the base have length $2\sqrt{2}\cos\theta$ and $2\sqrt{2}\sin\theta$ for projection onto a vertical plane; this forces $\theta = \pm\pi/4$, and any projection other than horizontal would increase these lengths.) Therefore Fig. 2b illustrates the only possibility, and V is outside. Note that if the circle were replaced by a square, and AB were a diagonal of this new doorway, the example would not succeed. The lines AS and BT must lie in the square as the base passes through, and then their intersection V is also in the square. Therefore a more delicate counterexample is required.

We change the base, previously square, to an equilateral triangle of altitude 2. Thus the pyramid becomes a tetrahedron C, to be passed through a square opening of side $\sqrt{2}$. For convenience the opening is turned so that the widest part, a diagonal of length 2, is horizontal. Then the first step translates the set in the y-direction until half the base passes through, and the altitude OP lies exactly in the opening—covering the diagonal from $(0,0,0)$ to $(2,0,0)$. The vertex is then at a point V which is still on the side $y < 0$ and therefore not through the door. We now rotate the set, at first through an angle $\pi/6$ around the diagonal OP. This lowers the vertex to a point V' on the x-y plane; the tetrahedron is as illustrated in Fig. 3a. Then a second rotation, through $\pi/4$ around the vertical axis, brings the vertex to the point $V'' = (4\sqrt{2}/3, 0, 0)$ in the doorway. At this stage the rest of the tetrahedron can pass directly through, by a final translation in the y-direction. The positions of the vertices are given by

$$\begin{array}{lcccccc}
A = \left(0, \ \dfrac{2}{\sqrt{3}}, \ 0\right) & & \left(0, \ 1, \ \dfrac{1}{\sqrt{3}}\right) & & \left(-\dfrac{1}{\sqrt{2}}, \ \dfrac{1}{\sqrt{2}}, \ \dfrac{1}{\sqrt{3}}\right) = A'' \\[2ex]
B = \left(0, \ -\dfrac{2}{\sqrt{3}}, \ 0\right) & \xrightarrow[\pi/6]{} & \left(0, \ -1, \ -\dfrac{1}{\sqrt{3}}\right) & \xrightarrow[\pi/4]{} & \left(\dfrac{1}{\sqrt{2}}, \ -\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{3}}\right) = B'' \\[2ex]
P = (2, \ 0, \ 0) & & (2, \ 0, \ 0) & & (\sqrt{2}, \ \sqrt{2}, \ 0) = P'' \\[2ex]
V = \left(\dfrac{4}{3}, -\dfrac{2}{\sqrt{3}}, \ \dfrac{2}{3}\right) & & \left(\dfrac{4}{3}, -\dfrac{4}{3}, \ 0\right) & & \left(\dfrac{4\sqrt{2}}{3}, \ 0, \ 0\right) = V''
\end{array}$$

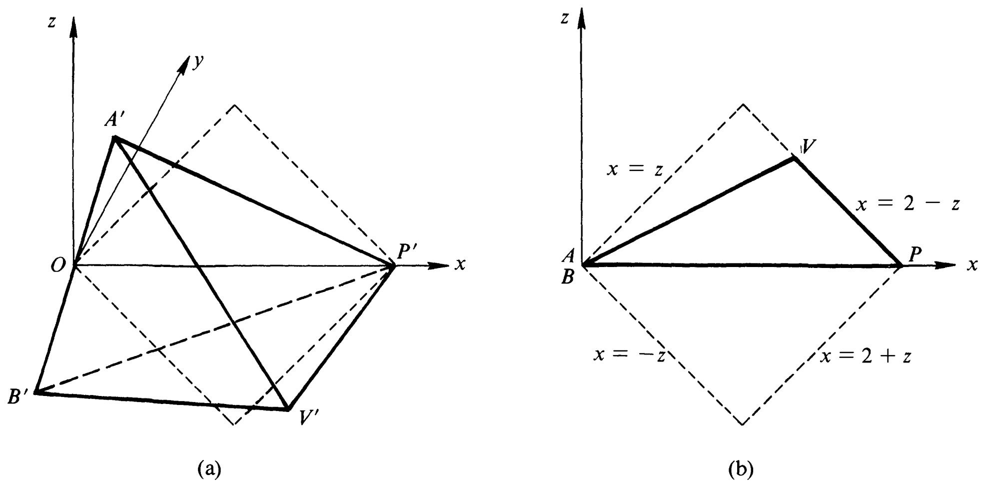


FIG. 3. (a) After one rotation. (b) Before rotation: front view.

We verify below that at every step the set C intersects the plane $y = 0$ within the given square $|z| \leq x \leq 2 - |z|$. In fact the projection of the original tetrahedron $ABPV$ *also* lies in this square, so that the set can pass through the door by pure translation (Fig. 3b). The point V barely fits, since its x and z coordinates add exactly to 2. As it stands, this set fails to establish that rotation may be required. Therefore we stretch the coordinates of A, B, and V by a factor $1 + \varepsilon$, so that the projection of V is shifted outside the square, and the base ABP (now isosceles) has just one altitude OP of length 2. The other altitudes are longer, and we can show that, as before, only projection in the y-direction fits the base into the doorway. The projection of $(1 + \varepsilon)V$ lies outside but the rotations through $\pi/6$ and $\pi/4$ now play their part; they move every point except O and P safely inside the door. Therefore the slightly larger tetrahedron is a convex set that goes through but not straight through.

We now compute the intersection of the tetrahedron, originally at $ABPV$, and the plane $y = 0$. The edge AB intersects at the origin O, and this remains true throughout the rotations. The edge AV intersects at its midpoint $M = (\frac{2}{3}, 0, \frac{1}{3})$, and since these coordinates satisfy $|z| < x < 2 - |z|$, M lies inside the square. Therefore the intersection is the triangle OMP (and there was no difficulty in translating the set to the position $ABPV$, with half the base through the door). During the first rotation, A and V move to

$$A_\phi = \left(0, \frac{2}{\sqrt{3}}c, \frac{2}{\sqrt{3}}s\right) \quad \text{and} \quad V_\phi = \left(\frac{4}{3}, -\frac{2}{\sqrt{3}}c - \frac{2}{3}s, -\frac{2}{\sqrt{3}}s + \frac{2}{3}c\right),$$

where c and s denote $\cos\phi$ and $\sin\phi$. The intersection with $y = 0$ is found at

$$M_\phi = \left(6c + \sqrt{3}s\right)^{-1}(4c, 0, 2);$$

since $c > 1/2$ for $0 \leq \phi \leq \pi/6$, this remains in the square and reaches $M' = (4/7, 0, 4/7\sqrt{3})$ at the end of this first rotation.

During the second rotation, which starts from $A'B'P'V'$ and turns around the z-axis, we have three edges to watch—the continued movement of $A'V'$ and also the edges which are initially at $B'P'$ and $P'V'$. The vertices are found, with $c = \cos\theta$, $s = \sin\theta$, $0 \leq \theta \leq \pi/4$, at

$$A_\theta = \left(-s, c, \frac{1}{\sqrt{3}}\right) \qquad P_\theta = (2c, 2s, 0)$$

$$B_\theta = \left(s, -c, -\frac{1}{\sqrt{3}}\right) \qquad V_\theta = \left(\frac{4}{3}(s + c), \frac{4}{3}(s - c), 0\right).$$

The edge $P_\theta V_\theta$ is the easiest; it stays in the plane $z = 0$, and its first component moves in from $x = 2$ at $\theta = 0$ to $x = 4\sqrt{2}/3$ at $\theta = \pi/4$. The edge $B_\theta P_\theta$ intersects $y = 0$ at

$$N_\theta = \frac{2sB_\theta + cP_\theta}{2s + c} = \frac{\left(2, 0, \frac{-2s}{\sqrt{3}}\right)}{2s + c},$$

which moves well inside the square $|z| \leqslant x \leqslant 2 - |z|$. Finally, the same is true for $A_\theta V_\theta$, which intersects at

$$M_\theta = \frac{\frac{4}{3}(c - s)A_\theta + cV_\theta}{\frac{4}{3}(c - s) + c} = \frac{\left(\frac{4}{3}, 0, \frac{4(c - s)}{3\sqrt{3}}\right)}{\frac{7}{3}c - \frac{4}{3}s}.$$

The sum $x + z$ decreases from its value at $\theta = 0$, where $M_\theta = M'$; thus it stays below 2. At the end of this second rotation, B'' is the only vertex still on the wrong side $y < 0$ of the door. But its projection is found at $x = 1/\sqrt{2}$, $z = -1/\sqrt{3}$, which lies in the square. Therefore all that remains of the tetrahedron—namely, a pyramid with vertex at B'' and base determined by 0, V'', and the final values $M'' = M_{\pi/4}$ and $N'' = N_{\pi/4}$—is moved through by translation in the y-direction.

We conclude that, even when a convex set can go through a door, it may be impossible to pack it into a rectangular box which does so.

Remark. Even for a box that does fit into a square doorway, the orientation that succeeds can change abruptly. Suppose the cross section of a large box is a rectangle of width w and height $h \leqslant w$, and the square still has sides of length $\sqrt{2}$. Then if $w \leqslant \sqrt{2}$, the rectangle can be aligned parallel to the square; if on the other hand $w + h \leqslant 2$, the rectangle fits after rotation through $\pi/4$. (The exact requirement is $w\cos\theta + h\sin\theta \leqslant 2$, if the rectangle is rotated through θ. But no intermediate position can give an improvement, since $w > \sqrt{2}$ and $w + h > 2$ imply

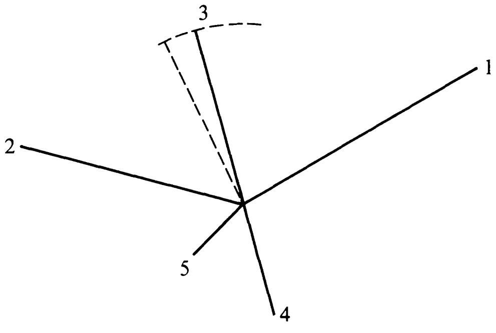
$$w\cos\theta + h\sin\theta > \sqrt{2}\cos\theta + \left(2 - \sqrt{2}\right)\sin\theta > \sqrt{2}$$

over the full range $0 < \theta < \pi/4$.) There is a critical ratio $h/w = \sqrt{2} - 1$ at which both $w = \sqrt{2}$ and $w + h = 2$ and the orientation of the largest possible box jumps from $\theta = 0$ to $\theta = \pi/4$.

3. Nonconvex Sets. For nonconvex sets we return to two dimensions. As it stands, the rule expressed in (1) prevents a large loop (a large circle, for example, with a piece removed) from passing through I. When the head starts through from $x < 0$, the tail is still on the wrong side $x > 0$. It seems reasonable to modify the rule; we can think instead of fixing the set C and moving the door. On the other hand we can make the passage more difficult by changing from a door to a hallway. Even in two dimensions it is an open question [**2**] to determine the largest set which can be moved around a corner in a hallway. Note that with convexity a straight hall is not difficult—by Theorem 1 the set can still go through with pure translation—but a nonconvex set might be able to enter the door $x = 0$, $|y| \leqslant 1$, when it could not travel down the hall bounded by $0 \leqslant x \leqslant L$, $y = \pm 1$.

To start on the general case, we consider the antithesis of a convex set. Let C be a union of finitely many line segments in R^2, each with an endpoint at the origin (Fig. 4a). If their lengths are ordered by $l_1, l_2, \ldots$, then it is l_3 that determines the shortest interval which can be passed around the set C. Moving the interval instead of the set, it encloses the longest segment of length l_1 as it comes toward the origin and the second longest as it leaves. Close to the origin the interval passes around the others, and it barely passes the segment of length l_3.

More generally, if there are several nodes and C is a connected union of line segments (the projection of an aluminum chair), then an algorithm to find the smallest interval is certainly



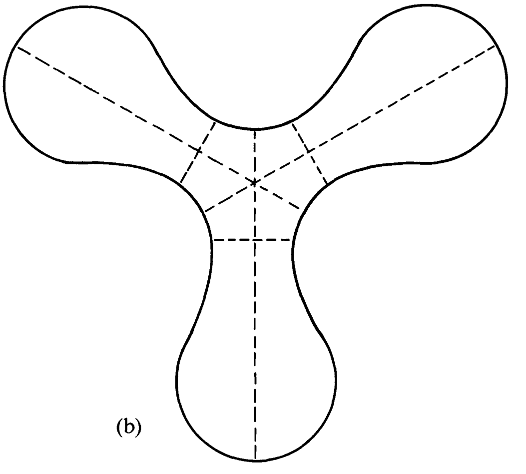


FIG. 4. Nonconvex sets.

possible. The same will be true for the simplest model of a three-dimensional chair—a box with line segments of equal length for legs. Our earlier remark suggests that the optimal motion will not depend smoothly on the dimensions of the box. And we mention also a more difficult problem: how to pass the human body between parallel bars.* In this case more than rigid motions are allowed, but Theorem 2 is still relevant. It is not clear, for a rectangular opening, when the passage of the head is the decisive step.

For the real problem, when C is a compact nonconvex set, we close with a conjecture. In R^2 we consider the *perpendicular chords* of C, those which are orthogonal at both ends to the boundary ∂C (assumed smooth, as in Fig. 4b). We believe that one of these chords (but not always the shortest) will determine the smallest interval which can pass around C. Its length will be the *width of the set*. In the convex case the width d is given by the shortest perpendicular chord and the conjecture is verified by Theorem 1.

We gratefully acknowledge the support of the National Science Foundation (MCS81-02371) and the Army Research Office (DAAG-29-80-K0033).

References

1. F. A. Valentine, Convex Sets, McGraw-Hill, New York, 1964.
2. L. Moser, Problem 66-11, SIAM Rev., 8 (1966) 381; see also SIAM Rev., 11 (1969) 75–78, SIAM Rev., 12 (1970) 582–586, this MONTHLY, 83 (1976) 188–189, and this MONTHLY, 84 (1977) 807–815.

*An ancient version of the Prisoner's Dilemma.

MISCELLANEA

81. It has been a fortunate fact in the modern history of physical science that the scientist constructing a new theoretical system has nearly always found that the mathematics ... required ... had already been worked out by pure mathematicians for their own amusement. ... The moral for statesmen would seem to be that, for proper scientific "planning," pure mathematicians should be endowed fifty years ahead of scientists.—R. B. Braithwaite, *Scientific Explanation*, Cambridge University Press, 1953, pp. 48–49.

BLOWING UP SINGULARITIES IN CLASSICAL MECHANICAL SYSTEMS

ROBERT L. DEVANEY*
Department of Mathematics, Boston University, Boston, MA 02215

Among smooth dynamical systems, some of the most complicated arise from Newton's equations in classical mechanics. Some of these systems (including many of those studied in elementary physics courses) are integrable, and hence their phase portraits are well understood. But there are many other nonintegrable systems whose phase portraits are still far from being completely understood. Examples of these systems include the well-known n-body problem of celestial mechanics when $n > 2$. For these systems, the set of time-periodic solutions of the differential equation is typically dense in the phase space of the system, but there are also many other much more complicated types of solutions. These include recurrent solutions, which never close up but which return infinitely often to any prescribed neighborhood of their starting position. Sometimes these solutions wind densely about a torus; other times they fill out a dense subset of an open set in a surface of constant energy. In these cases it may be impossible to tell what type of solution will be generated by a given initial condition, no matter how accurate the analytic, qualitative or numerical techniques one uses.

To make matters worse, specific Hamiltonian systems which arise in applications often suffer singularities as well. By a *singularity* we mean a point where the differential equation itself is undefined. A typical example of a singularity is a collision between two or more of the point masses in the Newtonian n-body problem. At collision, the differential equation breaks down: the velocities of the particles involved become undefined. A singularity or collision can create havoc among nearby solution curves. Solutions which pass near a singularity may behave in an erratic or unstable manner, and solutions which start out close to one another can end up far apart after passing by a singularity.

At the other end of the spectrum in dynamical systems are the gradient-like Morse-Smale systems. These systems feature none of the complicated solution curves that appear in Hamiltonian systems. There are never any periodic or recurrent solutions. The only "interesting" solutions are the rest points or equilibrium solutions, and each of these is simple in character; either it is a sink, a source, or a saddle point. Given a particular differential equation of this type, these rest points may often be found exactly, and their character determined explicitly. Moreover, all other solutions tend toward and away from one of these rest points. So the dynamics of a gradient-like Morse-Smale system are exceedingly simple and one can usually take such a system as being (qualitatively, at least) completely understood.

There would appear to be very little connection between complicated Hamiltonian systems and gradient-like Morse-Smale systems. Besides the above differences, the flow of a Hamiltonian system preserves a smooth volume in phase space, whereas gradient-like systems possess sinks, which necessarily rules out volume preservation. Also, Hamiltonian systems usually model conservative phenomena whereas gradient-like systems model dissipative phenomena.

Nevertheless, there is a place for gradient-like Morse-Smale systems in Hamiltonian mechanics. Using a technique due to McGehee [1974], one may make a simple change of variables which removes the singularity in a Hamiltonian system and which replaces it by a smooth surface or boundary called the *collision surface*. After a change of time scale, the transformed system extends

The author received his Ph.D. from the University of California, Berkeley, under the direction of Stephen Smale. He has taught at Northwestern University, Tufts University, and the University of Maryland. He is currently Associate Professor of Mathematics at Boston University. His research interests center primarily in dynamical systems, with special emphasis on classical mechanics, and his nonresearch interests include a fanatical addiction to opera.

*Partially supported by NSF Grant MCS 81-01855.

over this surface, and, amazingly enough, the system usually degenerates to a gradient-like Morse-Smale system on the boundary. Since the dynamics on the collision surface are therefore quite simple, one can use this information to investigate how solution curves behave near this surface, i.e., as they pass close to the singularity. Thus, one can understand completely the erratic and unstable behavior generated by a singularity in the equations of motion.

The aim of this paper is twofold. First, we will describe the technique of blowing up the singularity. Secondly, we will illustrate its usefulness by describing the behavior of several important classical mechanical systems, namely the Kepler problem and the anisotropic Kepler problem.

1. Near Collision Orbits of the Kepler and Anisotropic Kepler Problems. Let us begin by describing the type of qualitative behavior of solutions we will be able to understand using the technique of blowing up the singularity. Recall first the setting of the Newtonian Central Force or Kepler problem. A particle in the plane is attracted to the origin by a force which is inversely proportional to the square of the distance to the origin. Assuming that the mass and gravitational constants are normalized to one, Newton's Law gives the differential equation

$$q'' = -\left(\frac{1}{|q|^2}\right)\frac{q}{|q|} \tag{1.1}$$

where q is a vector in the plane.

The solutions of this second-order differential equation are well known from calculus or mechanics courses: all solutions lie on (possibly degenerate) conic sections.

We will be most concerned with collision and near-collision orbits. Collision orbits arise as follows. Suppose the particle is started from rest at some point q_0 in the plane. Then the particle travels along a straight line directly to the origin. The particle's velocity increases, and the velocity becomes infinite precisely when the particle reaches or collides with the origin.

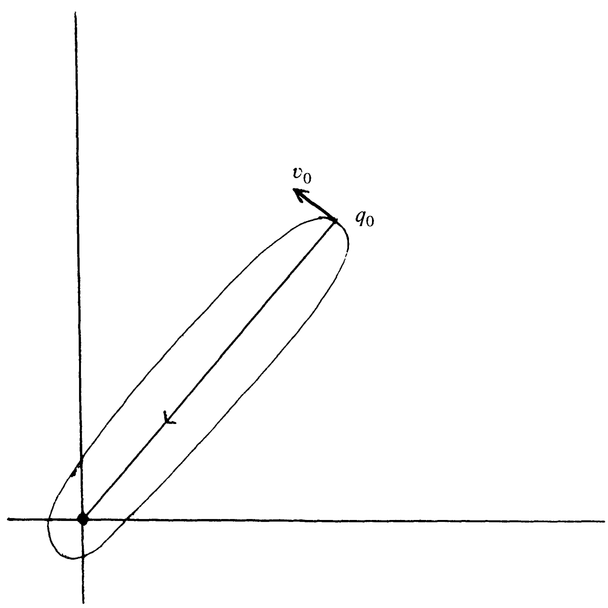


FIG. 1. A near collision orbit in the Kepler problem.

One also has similar ejection orbits. These emanate from the origin with infinite velocity and travel along rays with ever-decreasing velocity. If the particle eventually achieves zero velocity, then it turns around and retreats to the origin, thus giving a collision/ejection orbit. The question

we pose is: how does the existence of such collision/ejection orbits influence the dynamical behavior of nearby solution curves.

For the Kepler problem, the answer to this question is well known. Suppose we consider the original initial condition q_0 where the particle started from rest and eventually collided with the origin. This time, however, let us add a small component of velocity v_0 normal to the vector q_0. If this perturbation is small enough, the resulting solution curve lies on an ellipse, one of whose foci is at the origin. A sketch of such an orbit is shown in Fig. 1.

The line through the origin represents a collision/ejection orbit having zero velocity at q_0. Nearby solutions tend to wind around the origin exactly 360° before returning to their starting positions. (The fact that these nearby solutions close up is a special property of the Kepler problem which does not occur in most other mechanical systems.)

Now let us consider the anisotropic Kepler problem. This differential equation looks similar to that of the Kepler problem, but the orbit structure of this system is dramatically different.

This problem has become important recently in physics, where it serves as a nontrivial limiting classical mechanical system for certain quantum mechanical systems. See, for example, Gutzwiller [1973].

The differential equations for the anisotropic Kepler problem are

$$q_1'' = \frac{-\mu q_1}{\left(\mu q_1^2 + q_2^2\right)^{3/2}}$$

$$q_2'' = \frac{-q_2}{\left(\mu q_1^2 + q_2^2\right)^{3/2}} \tag{1.2}$$

where $q = (q_1, q_2)$ is a vector in the plane. Note that this system differs from the Kepler problem only in the additional parameter μ, which we take to be greater than one.

We may interpret this problem physically as the motion of a particle under the influence of a force at the origin, only this time the force is not directed toward the origin, but rather more toward the q_2-axis than the q_1-axis. This anisotropy has several immediate effects. For one thing, particles which start from rest do not necessarily fall directly into the origin. Only if the particle is started from rest on one of the coordinate axes does symmetry prevail and force a collision/ejection orbit. More importantly, other solutions behave dramatically differently from the Kepler problem, provided we have sufficient anisotropy. For this problem, sufficient anisotropy means $\mu > 9/8$. One of the goals of this paper is to explain where this number comes from and what erratic behavior ensues for $\mu > 9/8$.

For now, we content ourselves with sketching several typical solutions of the anisotropic problem when $\mu > 9/8$. If we perturb the collision/ejection orbits by adding a small amount of velocity in a direction normal to the q_1-axis, then the resulting solutions behave as in Figs. 2 and 3.

Note that the near-collision solutions differ tremendously from their counterparts in the Kepler problem. After passing close to collision with the origin, the particle circles 270° about the origin and then oscillates violently up or down the q_2-axis. We can find solution curves of this system which cross the q_2-axis an arbitrarily large number of times after passing around the origin.

This behavior is quite random or erratic: in fact, an important conjecture of Gutzwiller [1973] states that, for $\mu > 9/8$, the anisotropic Kepler problem is an ergodic system (on fixed negative energy levels). Behavior as illustrated in Figs. 2 and 3 is one of the pieces of evidence supporting this conjecture.

2. Hamiltonian Systems. Motivated by our examples of specific classical mechanical systems, we will take a rather limited view of Hamiltonian systems. Hamiltonian mechanics may be set in a much broader framework. See, for example, the books of Abraham and Marsden [1978] or Arnold [1978] for a more general treatment.

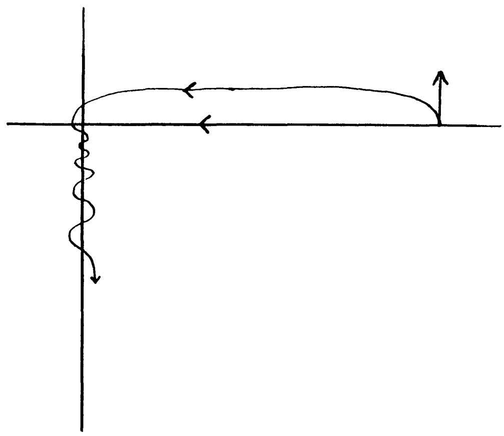


FIG. 2

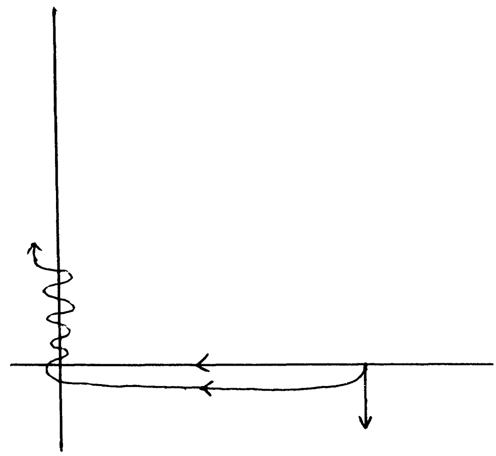


FIG. 3

Newton's equations can be written as

$$Mq'' = -\nabla V(q). \tag{2.1}$$

Here M is the mass matrix which is usually taken to be a diagonal matrix with positive entries $m_1, \ldots, m_n$. The vector $q = (q_1, \ldots, q_n)$ is an n-vector, and the real-valued function V is assumed sufficiently smooth. In light of our examples below, we will make several additional assumptions, all of which may be easily eliminated. First, we will assume that the system has two degrees of freedom, i.e., that q is a vector in the plane. Secondly, we will assume that V is homogeneous of degree -1; that is, $V(\lambda q) = \lambda^{-1} V(q)$. This implies that the function V is undefined at the origin, and therefore the differential equation (2.1) has a singularity at $q = 0$. We will assume that this singularity is isolated in $\mathbb{R}^2$. Finally, we will assume that the masses have been normalized to one, so that $M = I$.

All of these restrictions are easily eliminated. For example, the changes of coordinates below may be extended easily to higher dimensions, and they may be modified for homogeneous potentials of other degrees. See, for example, Devaney [1981].

Newton's equations are more conveniently written as a system of first order differential equations by introducing the velocity vector $p = q'$. Then (2.1) may be written

$$q' = p, \quad p' = -\nabla V(q) \tag{2.2}$$

recalling that we have assumed that $M = I$.

This system of first-order differential equations is a simple example of a Hamiltonian system. By this we mean the following: Consider the real-valued function

$$H(q, p) = \frac{1}{2}|p|^2 + V(q). \tag{2.3}$$

H is called the *Hamiltonian* or *total energy* function. Using H, (2.2) may be written in the form

$$q_i' = \frac{\partial H}{\partial p_i} \qquad p_i' = -\frac{\partial H}{\partial q_i} \tag{2.4}$$

for $i = 1, 2$. Equations in this form are called *Hamiltonian systems*. These systems enjoy many special properties. Perhaps the most important of these is conservation of total energy or, equivalently, the fact that H is constant along solution curves of (2.4). Indeed, this may be seen easily using the chain rule:

$$\frac{dH}{dt} = \sum_{i=1}^{2} \frac{\partial H}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial H}{\partial p_i}\frac{dp_i}{dt} = 0. \tag{2.5}$$

The term $\frac{1}{2}|p|^2$ in (2.3) is called the *kinetic energy* of the system and the function $V(q)$ is called the *potential energy*. Let Q denote the punctured plane $\mathbb{R}^2 - 0$. Q is called the *configuration space* of the system, while $Q \times \mathbb{R}^2$ (q, p-space) is the *phase space*.

Since (2.2) is a first-order system of differential equations, we think geometrically of (2.2) as defining a vector field on the space $Q \times \mathbb{R}^2$. Solutions of (2.2) may be viewed as parametrized curves which are everywhere tangent to the vector field. These curves are called *orbits* of the vector field, and our goal is to understand the qualitative behavior of these orbits near $q = 0$. Note that orbits give more information than just the position or configuration of the system in Q; the orbit also provides the velocity of the system at time t.

By the Existence and Uniqueness Theorem for ordinary differential equations, there is a unique orbit passing through each point in $Q \times \mathbb{R}^2$ starting at time t_0. Translations in time do not alter the solution curve passing through a given point in $Q \times \mathbb{R}^2$. However, not all orbits need be defined for all time. Certain orbits of the system run into the singularity as time either increases or decreases. An orbit is called a *collision orbit* if $q(t) \to 0$ as t increases to a finite limit. Similarly, an orbit is called an *ejection orbit* if $q(t) \to 0$ as t decreases to a finite limit. Generally, the velocities $p(t)$ become infinite at the moment of collision or ejection, so these orbits cannot be continued in any obvious way through the singularity.

Since H is constant along each solution curve, we may reduce the dimension of the system by considering only those solutions which satisfy $H(q, p) = e$. By (2.5), the vector field (2.2) is everywhere tangent to the level set $H^{-1}(e)$, and so solutions which begin on $H^{-1}(e)$ remain for all time on this set. That is, they satisfy for all time the *energy relation*

$$\frac{1}{2}|p(t)|^2 + V(q(t)) = e. \tag{2.6}$$

In general, $H^{-1}(e)$ is a smooth hypersurface in $Q \times \mathbb{R}^2$. $H^{-1}(e)$ is called an *energy surface*.

3. McGehee Coordinates. In this section we will introduce a simple change of coordinates due to R. McGehee [1974]. In this new coordinate system, after a change of time scale, we will be able to "read off" the behavior of solutions which pass close enough to $q = 0$.

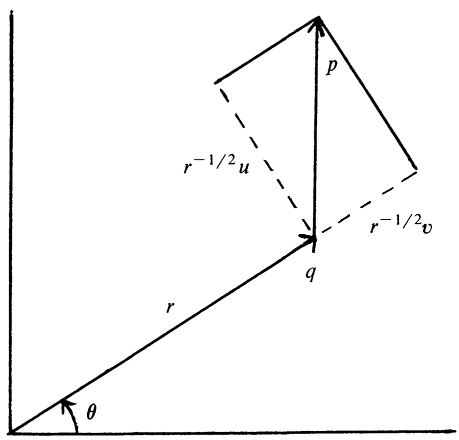


FIG. 4. McGehee coordinates.

Quite simply, the new coordinates we choose are polar coordinates in configuration space, together with scaled versions of the radial and angular components of velocity. See Fig 4. More precisely, let r and θ denote the usual polar coordinates on Q, and define v and u by

$$r^{-1/2}v = p \cdot (\cos\theta, \sin\theta) = \frac{(q \cdot p)}{r}$$
$$r^{-1/2}u = p \cdot (-\sin\theta, \cos\theta). \tag{3.1}$$

We call the coordinates (r, θ, v, u) McGehee coordinates on $Q \times \mathbb{R}^2$. In these new coordinates, one may check easily that the system (2.2) goes over to

$$\begin{aligned} r' &= r^{-1/2}v \\ \theta' &= r^{-3/2}u \\ v' &= r^{-3/2}\left(u^2 + \frac{1}{2}v^2 + V(\theta)\right) \\ u' &= r^{-3/2}\left(-\frac{1}{2}vu - V'(\theta)\right). \end{aligned} \tag{3.2}$$

Here $V(\theta)$ means the restriction of the potential energy to the unit circle in configuration space. Since the only singularity for V occurs at $q = 0$, it follows that both $V(\theta)$ and $V'(\theta)$ are smooth functions without singularities. The only singularity of (3.2) occurs at $r = 0$.

These formulas are easy to verify provided one recalls Euler's Formula for the gradient of a homogeneous function of degree $-k$:

$$q \cdot \nabla V(q) = -kV(q) \tag{3.3}$$

so that, in our case

$$q \cdot \nabla V(q) = -\frac{1}{r}V(\theta). \tag{3.4}$$

Also, since $V(q)$ is homogeneous of degree -1, $\nabla V(q)$ is homogeneous of degree -2, and, in particular

$$\nabla V(q) = \frac{1}{r^2}\nabla V(\cos\theta, \sin\theta). \tag{3.5}$$

The energy relation (2.6) is transformed in McGehee coordinates to

$$\frac{1}{2}(v^2 + u^2) + V(\theta) = re. \tag{3.6}$$

This new system may be viewed as a vector field on the space $(0, \infty) \times S^1 \times \mathbb{R}^2$, where S^1 represents the unit circle in Q. The singular set has been transformed into the boundary $r = 0$. But now, these singular points may be removed by a simple change of time scale

$$\frac{dt}{d\tau} = r^{3/2}. \tag{3.7}$$

That is, if we multiply the vector field given by (3.2) by the function $r^{3/2}$, the solution curves remain the same, but are parametrized differently. In the new time scale τ, the differential equation becomes

$$\begin{aligned} \dot{r} &= rv \\ \dot{\theta} &= u \\ \dot{v} &= u^2 + \frac{1}{2}v^2 + V(\theta) \\ \dot{u} &= -\frac{1}{2}vu - V'(\theta) \end{aligned} \tag{3.8}$$

where the dot indicates differentiation with respect to τ.

We will study this system in detail in a later section, but for now we list some of the most important properties of this vector field.

1. This is a smooth vector field on the space $[0, \infty) \times S^1 \times \mathbb{R}^2$; that is, the system (3.2) has now been extended over the boundary $r = 0$.
2. This boundary is invariant since $\dot{r} = 0$ when $r = 0$. This means that orbits of (3.8) which

begin with $r = 0$ remain for all time in the boundary $r = 0$.

3. The energy relation also extends to $r = 0$, giving

$$\frac{1}{2}(v^2 + u^2) + V(\theta) = 0. \tag{3.9}$$

The surface determined by this equation in $r = 0$ is also invariant and serves as a boundary for each energy surface. This surface is called the *collision surface* and is denoted by Λ.

4. Λ is independent of e.
5. The change of time scale (3.7) has the effect of slowing down collision orbits so that they now take infinitely long to reach collision. Orbits which previously passed close to collision now spend a long time near Λ. By continuity of solutions with respect to initial conditions, the behavior of such near-collision orbits can be determined from the qualitative behavior of solutions on Λ. In the next few sections we will show that these solutions are easy to understand; generically, (3.8) restricts to a gradient-like Morse-Smale system on Λ.

We finally remark that Λ and the orbits on Λ are totally fictitious in terms of the original system. These are not solutions of the original system. Rather, one may think of Λ as what one sees when one examines the boundary under a microscope.

4. The Kepler Problem. Before describing the nature of solutions on the collision surface, we return to the Kepler problem to show how McGehee coordinates allow us to "read off" the local behavior of near-collision orbits described in §1.

Recall that the differential equation for the Kepler problem is

$$q'' = -\left(\frac{1}{|q|^2}\right) \cdot \frac{q}{|q|}. \tag{4.1}$$

This may be written as the Hamiltonian system

$$\begin{aligned} q' &= p \\ p' &= -\nabla V(q) \end{aligned} \tag{4.2}$$

where the potential energy $V(q)$ is given by $-1/|q|$. In McGehee coordinates, we therefore have

$$\begin{aligned} \dot{r} &= rv \\ \dot{\theta} &= u \\ \dot{v} &= \frac{1}{2}v^2 + u^2 - 1 \\ \dot{u} &= -\frac{1}{2}vu \end{aligned} \tag{4.3}$$

with energy relation

$$\frac{1}{2}(v^2 + u^2) - 1 = re \tag{4.4}$$

since the potential energy is identically -1 on the unit circle in the plane. When $r = 0$, the energy relation shows that Λ is a cross-product of two circles, i.e., a torus given by

$$\begin{aligned} &\frac{1}{2}(v^2 + u^2) = 1 \\ &\theta \text{ arbitrary.} \end{aligned} \tag{4.5}$$

The vector field on Λ is given by

$$\dot{\theta} = u, \quad \dot{v} = \frac{1}{2}u^2, \quad \dot{u} = -\frac{1}{2}vu, \tag{4.6}$$

where we have used the energy relation to simplify $\dot{v}$. This system of equations is easy to understand.

We first note that $\dot{v} > 0$ provided $u \neq 0$. This means that the v-coordinate must increase along any orbit on Λ which stays away from $u = 0$. On the other hand, if $u = 0$, the right-hand side of (4.6) vanishes. Therefore, the constant function $\theta = \theta^*$, $u = 0$, and $v = \pm\sqrt{2}$ (determined by the energy relation) is a solution of (4.6) for any θ^*. These constant solutions are called *rest* or *equilibrium points* for the system.

We thus have two circles of rest points for (4.6), one given by θ arbitrary, $u = 0$, and $v = \sqrt{2}$ and the other by θ arbitrary, $u = 0$, and $v = -\sqrt{2}$. All other nonequilibrium solutions must travel from this second circle to the first, since v must increase along solution curves.

This proves that, on Λ, the system is gradient-like, for a *gradient-like vector field* is one for which there is a function which increases along all nonequilibrium solutions.

We can in fact say much more about the solutions on Λ. Suppose we introduce the angular variable ψ in each v, u-plane via

$$\begin{aligned} u &= \sqrt{2}\cos\psi \\ v &= \sqrt{2}\sin\psi. \end{aligned} \tag{4.7}$$

In θ, ψ-coordinates, one checks easily that (4.6) may be written

$$\begin{aligned} \dot{\theta} &= \sqrt{2}\cos\psi \\ \dot{\psi} &= \frac{1}{\sqrt{2}}\cos\psi. \end{aligned} \tag{4.8}$$

Eliminating time from this equation, we get the simple first order differential equation

$$\frac{d\psi}{d\theta} = \frac{1}{2}. \tag{4.9}$$

In the θ, ψ-plane, it follows that solutions lie on straight lines of slope $\frac{1}{2}$. These solutions are sketched in Fig. 5.

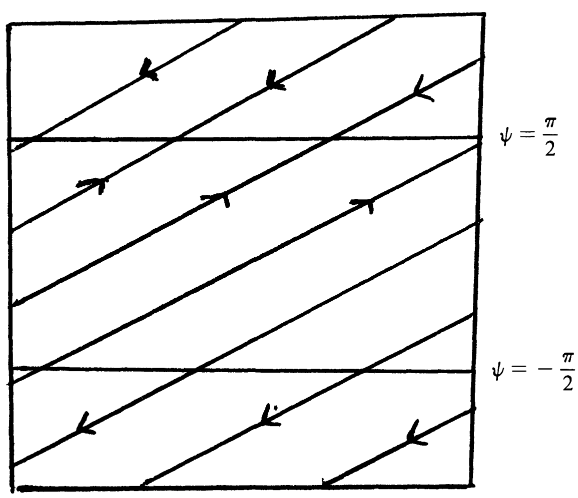


FIG. 5. The solution curves in the θ,ψ-plane.

Note that the two circles of rest points in these coordinates are given by $\psi = \pm\pi/2$. For each θ^*, there is a pair of solution curves which travel from the rest point at $\theta = \theta^*$, $\psi = -\pi/2$ up to the rest point at $\theta = \theta^*$, $\psi = \pi/2$. The curves leaving the rest point $\theta = \theta^*$, $\psi = -\pi/2$ are called the *unstable manifold* of this point, while the curves approaching $\theta = \theta^*$, $\psi = \pi/2$ are called the *stable manifold*. It is a coincidence (no pun intended) due to the very special nature of the Kepler problem that these stable and unstable manifolds coincide for each θ^*.

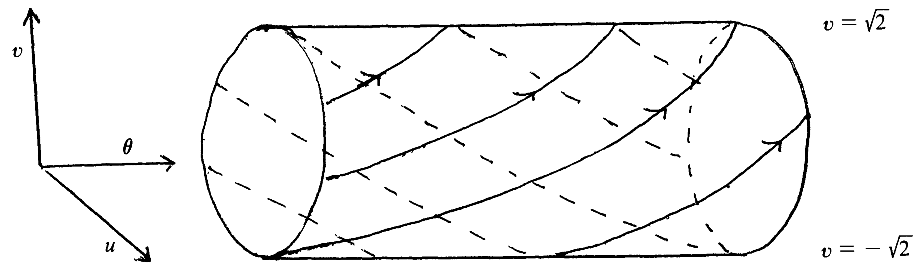


FIG. 6. Solution curves on Λ. Λ is a torus with $\theta = -\pi$ and $\theta = \pi$ identified.

In McGehee coordinates, the solution curves on Λ may be pictured as in Fig. 6.

Now recall the collision/ejection orbits described in §1. These solutions began and ended at the singularity. Since (4.3) is a smooth vector field, these solutions must therefore be asymptotic in both time directions (in the τ-time scale) to Λ. Also recall that these solutions were constrained to lie on a fixed ray, say $\theta = \theta^*$, with velocity always parallel to this ray. This means that $\theta = \theta^*$ and $u = 0$ identically along such solutions. Since the radial velocity is positive at ejection and negative at collision, it follows that this solution must travel from the rest point at $\theta = \theta^*$, $u = 0$, $v = \sqrt{2}$ to $\theta = \theta^*$, $u = 0$, $v = -\sqrt{2}$ in the space $r > 0$.

This orbit is hard to visualize in the four-dimensional (r, θ, v, u)-space, but it is drawn schematically in Fig. 7.

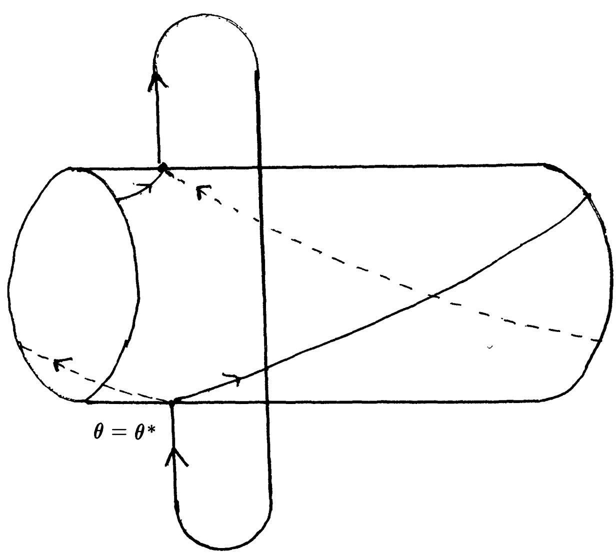


FIG. 7. A collision/ejection orbit.

Now let us return to the original question—how do near-collision orbits behave? The answer to this question is immediate from Fig. 7. Solutions near the collision orbit along $\theta = \theta^*$ may behave in three possible ways. The first possibility is that they begin and end in collision along a nearby ray. This corresponds to solutions tending to and from nearby rest points and is not the typical behavior. Most solutions miss the rest points on Λ and therefore, by continuity with respect to initial conditions, are forced to follow one of the two branches of the unstable manifold of the rest point at $\theta = \theta^*$, $v = -\sqrt{2}$, $u = 0$. Both of these branches tend ultimately to the rest point at $\theta = \theta^*$, $v = \sqrt{2}$, $u = 0$, however, so nearby solutions must come close to this rest point also. At this point, however, the ejection orbit at this point takes over and solutions following either

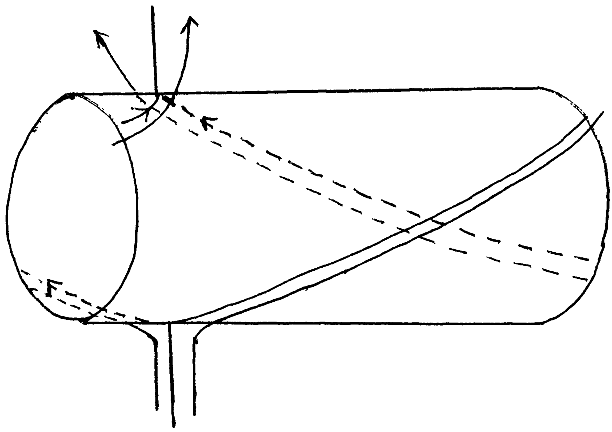


FIG. 8. Near-collision solutions of the Kepler problem near Λ.

possible branch must exit a neighborhood of Λ near the ray $\theta = \theta^*$. See Fig. 8.

These two possible behaviors are characterized by what happens to θ as the solution passes near the collision surface. In the first case, θ increases by 2π along the "front" branch of the unstable manifold, so θ does essentially the same for near-collision orbits. Similarly, in the second case, θ decreases by 2π along the "rear" branch of the unstable manifold, and so we expect the same behavior for near-collision orbits. Indeed, this is exactly what happens for near-collision orbits as we illustrate in Fig. 9.

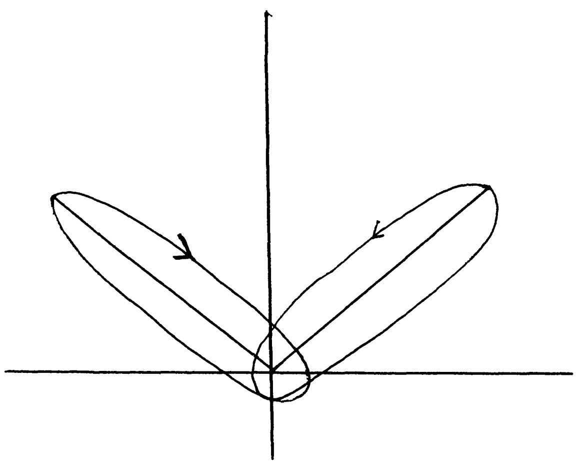


FIG. 9. Near-collision solutions in configuration space.

We remark that the fact that both near-collision orbits in Fig. 9 close up is a global phenomenon not predicted by the method above. However, this method does exclude any other *local* behavior near the singularity and, in the sense, does provide a complete qualitative picture of near-collision orbits.

We also remark that the system on Λ is not a Morse-Smale system and therefore is atypical. There are several reasons for this. First, rest points are not isolated in Λ. And secondly, the stable and unstable manifolds of rest points in Λ match up exactly. Both of these phenomena do not occur in Morse-Smale systems and neither occur generically when we "blow up" the singularity of a typical Hamiltonian system. This will be proved in the next section.

5. Solutions on the Collision Surface. In this section, we return to the general setting to describe the typical behavior of solutions on the collision surface. Recall that, in McGehee coordinates, the system on Λ is given by

$$\begin{aligned} \dot{\theta} &= u \\ \dot{v} &= u^2 + \frac{1}{2}v^2 + V(\theta) \\ \dot{u} &= -\frac{1}{2}vu - V(\theta) \end{aligned} \tag{5.1}$$

while Λ itself is determined by the equation

$$\frac{1}{2}(v^2 + u^2) + V(\theta) = 0. \tag{5.2}$$

One may prove easily that Λ is a smooth surface provided 0 is a regular value of $V(\theta)$, i.e., $V'(\theta_0) \neq 0$ whenever $V(\theta_0) = 0$. This happens for almost all potentials.

Using the energy relation, (5.1) may be written

$$\begin{aligned} \dot{\theta} &= u \\ \dot{v} &= \frac{1}{2}u^2 \\ \dot{u} &= -\frac{1}{2}vu - V'(\theta) \end{aligned} \tag{5.3}$$

As in the Kepler problem, it follows that $\dot{v} > 0$ provided $u \neq 0$. If $u = 0$, there are two possibilities. The first is $u = 0$, but $V'(\theta) \neq 0$. Since $\dot{u} \neq 0$, it then follows that the solution through this point immediately leaves the set where $u = 0$. Hence the coordinate v must increase along any solution which avoids the set of points with $u = 0$, and $V'(\theta) = 0$. But these are precisely the set of points where the right hand side of (5.3) vanishes. Therefore, we have shown Proposition 1.

PROPOSITION 1.

i. *The coordinate v increases along all nonequilibrium point orbits of* (5.3). *Therefore*, (5.3) *is a gradient-like system.*
ii. *The set of equilibrium points on Λ is given by all points satisfying*

$$\begin{aligned} &\text{(a)} \quad u = 0 \\ &\text{(b)} \quad V'(\theta_0) = 0 \\ &\text{(c)} \quad v = \pm\sqrt{-2V(\theta_0)}. \end{aligned}$$

Part ii(c) follows immediately from the energy relation (5.2). Thus rest points on Λ are in two-to-one correspondence with *critical points* of $V(\theta)$, i.e., points where V' vanishes.

Our goal now is to prove that (5.3) is usually a Morse-Smale system. To do this, we first need several definitions. The *linearization* of a system of differential equations is simply the Jacobian of the right-hand side of the system. We will be interested in the linearization at a rest point on Λ. One may check easily that the linearization of (5.3) at a rest point gives the matrix

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -V''(\theta) & 0 & -\frac{1}{2}v \end{pmatrix} \tag{5.4}$$

where $V'(\theta) = 0$ and $v = \pm\sqrt{-2V(\theta)}$. One of the eigenvalues of this matrix is always zero. The other two are called the *characteristic exponents* of the rest point. In our case, these exponents may

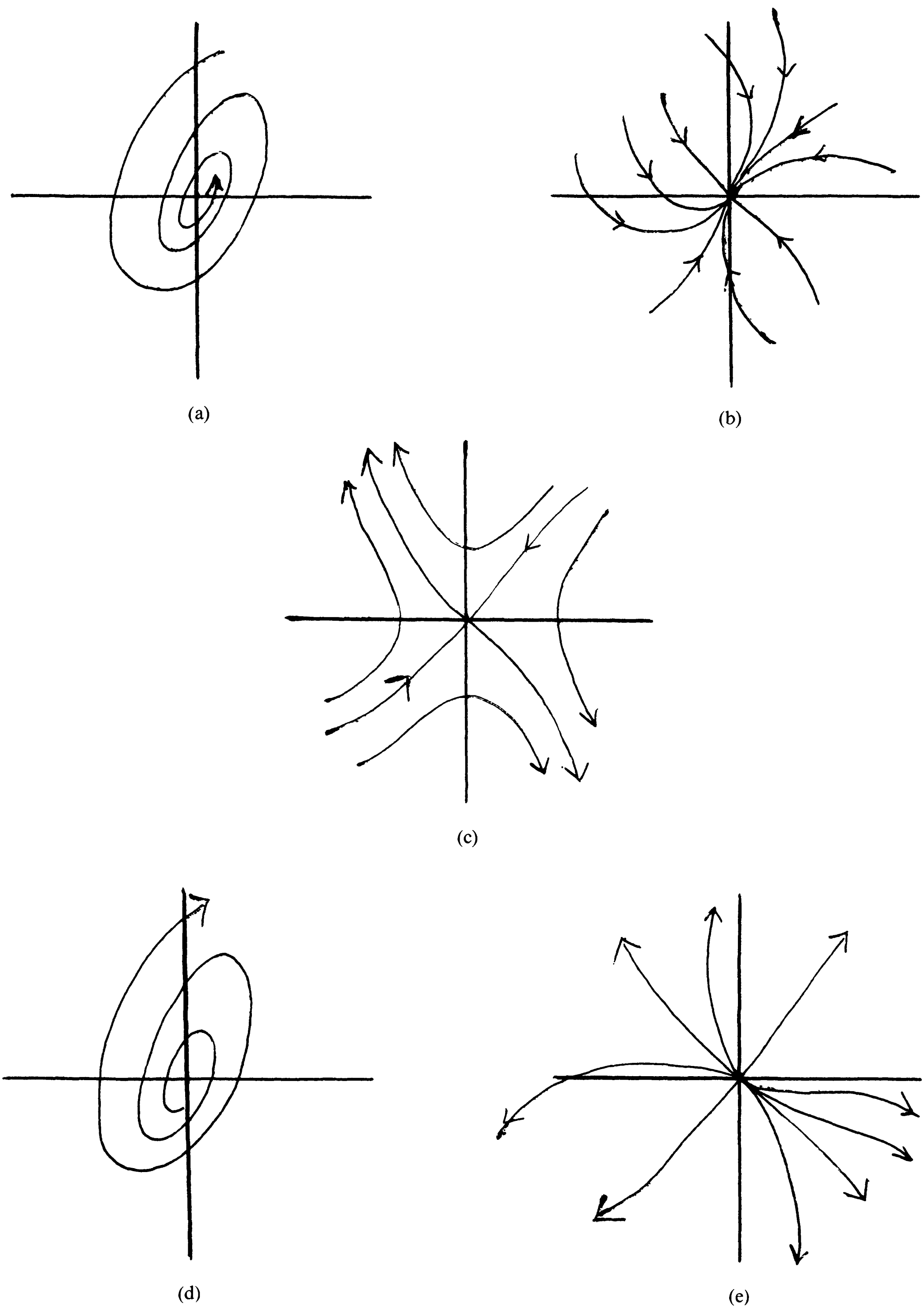


FIG. 10. a, b are sinks; c is a saddle; and d, e are sources.

be calculated easily and are given by

$$\zeta^{\pm} = -\frac{v}{4} \pm \frac{1}{4}\sqrt{v^2 - 16V''(\theta)}\,. \tag{5.5}$$

The importance of the characteristic exponents is that their real parts determine the qualitative behavior of nearby solutions in Λ. A rest point is called *hyperbolic* provided both characteristic exponents have nonzero real parts. Note that, in our case, the rest points are hyperbolic iff $V''(\theta) = 0$. Critical points of V which satisfy $V''(\theta) \neq 0$ are called *nondegenerate critical points*. Such critical points form the basic building blocks of Morse Theory (Milnor [1963]) and it is well known that almost all functions have the property that all critical points are nondegenerate (These functions are called Morse functions).

In our simple case, hyperbolic rest points come in three distinct types determined by the number of characteristic exponents with negative real parts. If both real parts are negative, the rest point is a *sink* and all nearby solutions tend toward the rest point. If both real parts are positive, the rest point is a source and nearby solutions tend away from the rest point. The mixed case, one positive and one negative, is called a *saddle* and features a pair of solutions tending asymptotically toward the rest point and a pair tending asymptotically away. Typical phase portraits for each of these cases are shown (in the plane) in Fig. 10. For more details, we refer the reader to the books of Hirsch-Smale [1974] or Arnold [1973]. For our purposes, the important point is we can determine the local behavior of solutions on Λ near a rest point provided we know the characteristic exponents. Since these exponents depend only on $V(\theta)$, they can usually be calculated quite easily.

The two solution curves emanating from the saddle point are called the *unstable manifold* of the rest point, while the two solutions tending toward the rest point are called the *stable manifold*. It follows that the stable and unstable manifolds of saddles in Λ form a union of curves in Λ. One can prove that, for most potential energy functions, V, no stable manifold matches up exactly with an unstable manifold in Λ. The proof involves perturbation techniques and is given in Devaney [1980].

Finally, a vector field is called gradient-like and Morse-Smale provided

i. There is a smooth function which increases along all non-equilibrium point orbits.
ii. All rest points are hyperbolic.
iii. No stable manifold of a saddle matches up with an unstable manifold of a saddle.

We have therefore proved in our simple case:

PROPOSITION 2. *For most potential energy functions, the vector field on Λ is gradient-like and Morse-Smale.*

As we noted earlier, in the Kepler problem, the vector field on Λ is not Morse-Smale as it violates conditions ii and iii above. However, our next example, the anisotropic Kepler problem, is Morse-Smale and will serve to illustrate how easily one may understand such systems.

6. The Anisotropic Kepler Problem. The goal of this section is to describe how the dynamics on the collision surface of a particular classical mechanical system determines easily and efficiently the (rather complicated) behavior of near-collision orbits. The system we will describe is the anisotropic Kepler problem. Recall the equations of motion for this system (1.2). They may be written in Hamiltonian form (2.2) using the potential energy function

$$V(q) = \frac{-1}{\sqrt{\mu q_1^2 + q_2^2}}\,. \tag{6.1}$$

This function is homogeneous of degree -1 and admits an isolated singularity at the origin. The parameter μ is chosen to be greater than one. When $\mu = 1$, we have the usual Kepler problem as

described in §4. When $\mu > 9/8$, the dynamics of the anisotropic problem are completely different from the Kepler problem, as we described in §1.

Restricted to the unit circle, the potential energy is given by

$$V(\theta) = \frac{-1}{\sqrt{\mu \cos^2\theta + \sin^2\theta}} \tag{6.2}$$

whose graph is sketched in Fig. 11.

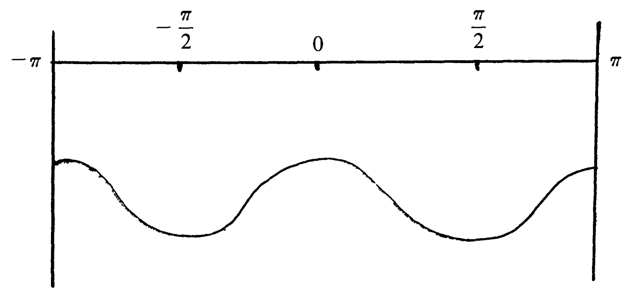


FIG. 11. The graph of $V(\theta)$ for the anisotropic Kepler problem.

One checks easily that $V(\theta)$ has nondegenerate maxima at $\theta = 0, \pi, 2\pi$, etc., and nondegenerate minima at $\theta = \pi/2, 3\pi/2$, etc. The collision surface is determined from the energy relation (3.9) and is sketched in Fig. 12.

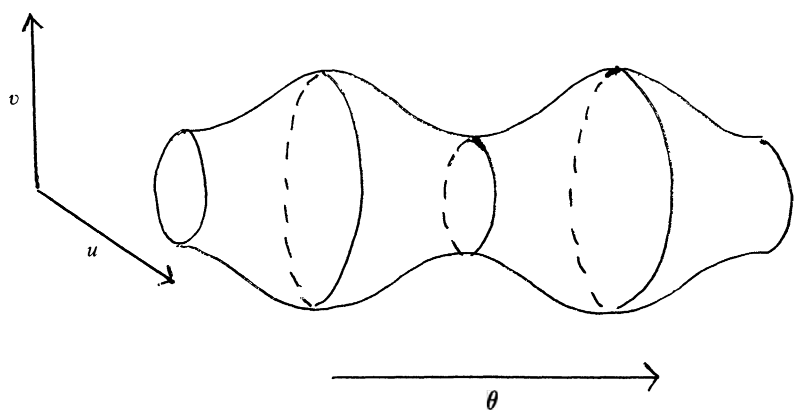


FIG. 12. The collision surface Λ for the anisotropic Kepler problem.

Note that Λ is a "bumpy" torus obtained by rotating the graph of $\sqrt{-2V(\theta)}$ around the θ-axis and identifying the circles at $\theta = 0$ and $\theta = 2\pi$.

By Proposition 1, there are eight rest points on Λ and (5.5) determines each of their characteristic exponents. We leave it to the reader to check that there are two sinks, two sources, and four saddles as depicted in Fig. 13. Indeed, since the v-coordinate must increase along nonequilibrium solutions, this must be the case.

We make one crucial additional observation. The characteristic exponents (5.5) in this case may be computed to be

$$\zeta^{\pm} = -\frac{1}{4}v \pm \frac{1}{2}\frac{\sqrt{9-8\mu}}{2} \tag{6.3}$$

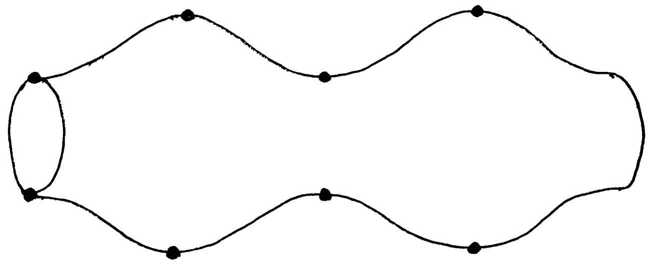


FIG. 13. The rest points on Λ.

when $\theta = \pi/2$ or $3\pi/2$. Hence these exponents are complex iff $\mu > 9/8$. According to Fig. 10, solutions therefore spiral into the sinks and away from the sources for these parameter values. When $1 < \mu < 9/8$, solutions tending to sinks or away from sources do not spiral, but rather behave as in Fig. 10b.

Gutzwiller [1973] has determined numerically the behavior of all stable and unstable manifolds of saddles. He finds that none of them coincide, so that the phase portrait of the vector field on Λ is as shown in Fig. 14.

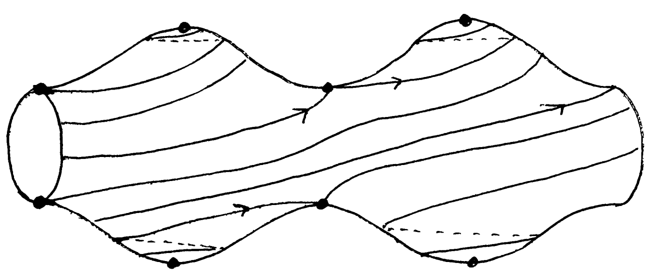


FIG. 14. Solution curves on Λ.

This may be verified by introducing an angular variable ψ in exactly the same way as we did for the Kepler problem. More precisely, let

$$\begin{aligned} u &= \sqrt{-2V(\theta)}\cos\psi \\ v &= \sqrt{-2V(\theta)}\sin\psi. \end{aligned} \tag{6.4}$$

Then the vector field on Λ reduces to

$$\begin{aligned} \dot{\theta} &= \sqrt{-2V(\theta)} \quad \cos\psi \\ \dot{\psi} &= \frac{\sqrt{-V(\theta)}}{\sqrt{2}} \quad \cos\psi + \frac{V'(\theta)\sin\psi}{\sqrt{-2V(\theta)}}. \end{aligned} \tag{6.5}$$

Eliminating time, one obtains the simpler first order equation

$$\frac{d\psi}{d\theta} = \frac{1}{2} - \frac{1}{2}\frac{V'(\theta)}{V(\theta)}\tan\psi. \tag{6.6}$$

Recall now from §1 that there are four collision/ejection orbits which remain for all time on the coordinate axes. These solutions therefore satisfy $\theta = 0$, $\pi/2$, π, $3\pi/2$ and $u = 0$ identically. Several of them may be visualized as in Fig. 15.

Note that each of these orbits emanates from a rest point with $v > 0$ and approaches a rest point with $v < 0$. This corresponds to positive radial velocity at ejection and negative radial velocity at collision.

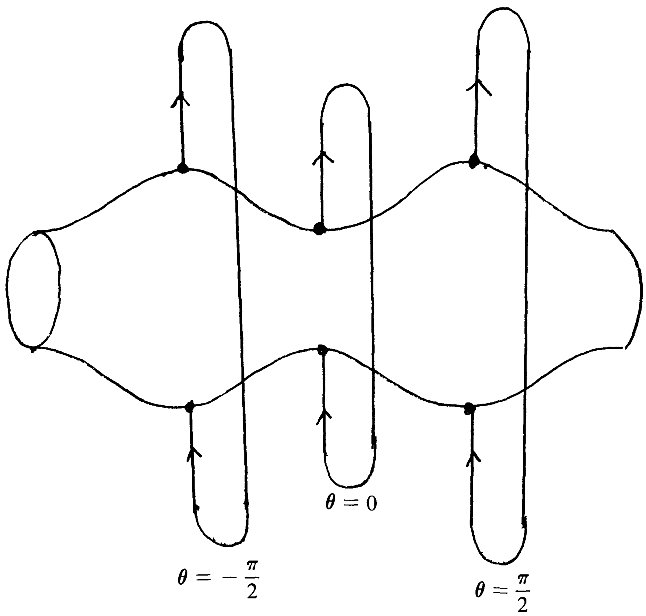


FIG. 15. Collision/ejection orbits.

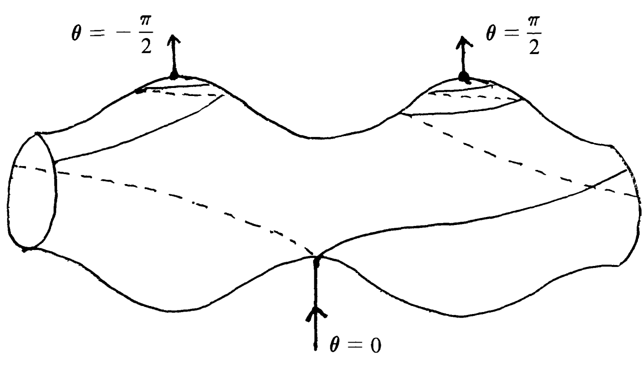


FIG. 16

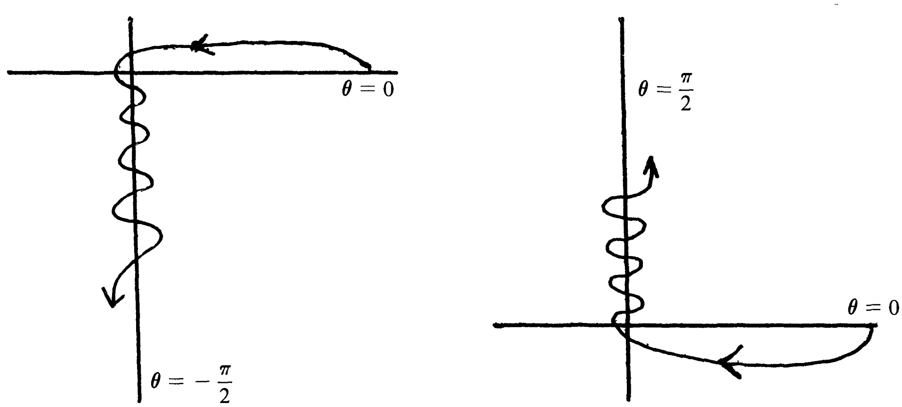


FIG. 17. Near-collision solutions in configuration space.

Let us consider in detail what happens to a solution near the collision/ejection orbit travelling along $\theta = 0$. As this orbit approaches the rest point with $v < 0$, it has two choices: either it may follow the branch of the unstable manifold of the saddle on the front face of Λ or else the branch on the back face. In the former case, the branch is asymptotic to the rest point at $\theta = 3\pi/2$, $v > 0$ after travelling $3\pi/2$ θ-units around Λ. In the latter case, the branch is asymptotic to the rest point at $\theta = -3\pi/2$ after travelling $-3\pi/2$ θ-units around Λ. See Fig. 16.

Nearby solutions must follow one of these two branches until coming close to the next rest point. Thereafter they follow the corresponding ejection orbit. Note that if $\mu > 9/8$, continuity of solutions with respect to initial conditions forces these solutions to spiral around this ejection orbit. In the r, θ configuration space, this corresponds to an oscillation about the $\theta = \pi/2$ or $\theta = 3\pi/2$ axis. Note that the closer we start to the original collision orbit, the more the orbit is forced to oscillate about the q_2-axis. This behavior is depicted in the original configuration space in Fig. 17.

7. Conclusion. The methods and results discussed in this paper may be applied to many more mechanical systems than those meeting our original restrictions. For example the restriction that V be homogeneous of degree -1 may be replaced by homogeneity of degree $-k$. The only changes are in the scaling factor (3.1) where we choose instead $r^{k/2}$ and in the change of time scale (3.7) where we choose

$$\frac{dt}{d\tau} = r^{1+k/2}.$$

As long as $k \neq 2$, one may prove easily that the system on Λ is gradient-like and Morse-Smale. See Devaney [1981] for details. The case $k = 2$ is special and does not fit into this general setting. A nice application of these techniques to potentials homogeneous of degree $-k$ is provided by McGehee [1982] who considers central force potentials of the form $V(q) = -1/|q|^k$.

Another important restriction to lift is that the mass matrix is the identity. This may be accomplished by working in a different metric on the plane where the dot product $\xi \cdot \eta$ is defined by

$$\xi \cdot \eta = \xi \cdot M\eta.$$

With this proviso, McGehee coordinates may be introduced as above with only slight changes in the resulting equations. See Devaney [1981].

Finally, we remark that many important classical mechanical systems besides the ones described in §1 may be treated by the methods above. Systems that arise in celestial mechanics are particularly suited to this process, for there singularities really are collisions between two or more of the masses. We mention briefly some of the versions of the classical n-body problem which have been discussed using these methods: the collinear three-body problem studied by McGehee [1974], the planar isosceles three-body problem treated by Devaney [1980], Lacomba and Losco [1980], Moeckel [1982], and Simo [1981], the trapezoidal four-body problem studied by Lacomba [1981], and finally the full-planar, three-body problem treated by Moeckel [1982]. Despite the amount of work already done on these problems, there remains much to be done in order to completely understand singularities in classical mechanical systems.

References

1. V. I. Arnold, Mathematical Methods of Classical Mechanics, Springer-Verlag, New York, Graduate Texts in Mathematics, vol. 60, 1978.

2. ______, Ordinary Differential Equations, M.I.T. Press, Cambridge, MA, 1973.

3. R. Abraham and J. Marsden, Foundations of Mechanics, Benjamin/Cummings, Reading, MA, 1978.

4. R. Devaney, Singularities in Classical Mechanical Systems, In Ergodic Theory and Dynamical Systems I, Birkhäuser, Boston, 1981, pp. 211–333.

5. ______, Morse-Smale singularities in simple mechanical systems, J. Differential Geom., 15 (1980) 285–305.

6. ______, Triple collision in the planar isosceles three body problem, Invent. Math., 60 (1980) 249–267.
7. ______, Collision orbits in the anisotropic Kepler problem, Invent. Math., 45 (1978) 221–251.
8. M. Gutzwiller, The anisotropic Kepler problem in two dimensions, J. Math. Phys., 14 (1973) 139–152.
9. ______, Bernoulli sequences and trajectories in the anisotropic Kepler problem, J. Math. Phys., 18 (1977) 806–823.
10. M. Hirsch and S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, New York 1974.
11. E. Lacomba, Quadruple collision in the trapezoidal four body problem, to appear in Classical Mechanics and Dynamical Systems, Marcel Dekker, New York.
12. E. Lacomba, and L. Losco, Triple collision in the isosceles three body problem, Bull. Amer. Math. Soc., 3 (1980) 710–714.
13. R. Moeckel, Orbits of the three body problem which pass infinitely close to triple collision. To appear.
14. ______, Orbits near triple collision in the three body problem. To appear.
15. R. McGehee, Triple collision in the collinear three body problem, Invent. Math., 27 (1974) 191–227.
16. ______, Double collisions for non-Newtonian potentials. To appear.
17. J. Milnor, Morse Theory, Princeton University Press. Annals of Math. Studies, no. 51, 1963.
18. C. Simo, Analysis of triple collision in the isosceles problem, In Classical Mechanics and Dynamical Systems, Marcel Dekker, New York, 1981.

PETRI NETS AND MARKED GRAPHS—MATHEMATICAL MODELS OF CONCURRENT COMPUTATION

RICHARD JOHNSONBAUGH*
Department of Mathematics, Chicago State University, Chicago, IL 60628

TADAO MURATA**
Department of Information Engineering, University of Illinois at Chicago, Chicago, IL 60680

Introduction. Petri nets and related graph models are recent tools used to model various classes of systems, especially systems involving parallel computations and concurrent processes.

Currently, one text on Petri nets is available [**44**] and Springer has published the proceedings of an advanced course on Petri nets held in Hamburg, West Germany, October 1979 [**5**]. Nevertheless, many of the results are still available only in conference proceedings and technical reports. It is hoped that this paper will introduce Petri nets to a wider audience and will give some of the flavor of the mathematics employed. Other expository papers are [**1**], [**33**], [**45**].

I. Definitions and History of Petri Nets. A Petri net is a particular kind of directed graph,

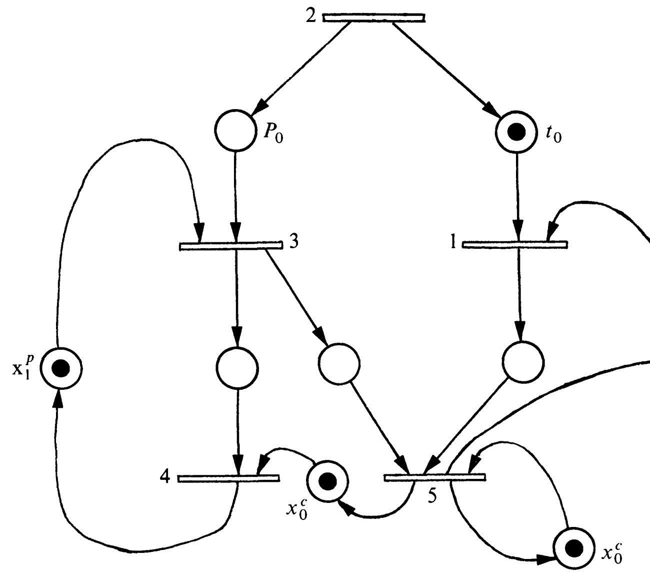
*Supported in part by NSF Grant MCS 80-09804.

Richard Johnsonbaugh received his Ph.D. degree in Mathematics from the University of Oregon in 1969. Since 1972 he has been with the Department of Mathematics at Chicago State University except for the 1978–79 academic year, which was spent at the University of Illinois at Chicago under an NSF Professional Development Grant. His research interests include applied graph theory, microcomputers, and discrete mathematics.

**Supported in part by NSF Grants ENG 78-05933 and ECS 81-05649.

Tadao Murata received his Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1966. He has been with the Department of Information Engineering at the University of Illinois at Chicago since 1966, except for occasional leaves of absence. From 1968–1970 he taught at Tokai University, Tokyo, Japan; and from 1976–1977 at the University of California at Berkeley. He was invited to visit Germany as a guest researcher in 1979. He was invited to visit France also as a guest researcher in 1981. He is currently serving as a member on the panel of the National Academy of Sciences, studying international developments in computer science. His research interests include concurrent computer systems, network theory, and applied graph theory.

together with an initial state called a *marking*. The graph has two kinds of vertices, *places* and *transitions*. Places are drawn as circles, and transitions are drawn as bars or rectangular boxes. Directed arcs may exist only between different types of vertices. (Such a graph is called a *bipartite graph*.) (See Fig. 1.)



Transition Firings Mean:
1: Compute $f(x_{n-1}^c, t_{n-1})$
2: Increment t_n
3: Compute $f(x_n^p, t_n)$
4: Compute Equation (1)
5: Compute Equation (2)

FIG. 1. Petri net representation of a parallel predictor-corrector algorithm.

A marking assigns to each place a nonnegative integer. If a marking assigns to place p the nonnegative integer n, we say that p has n *tokens*. Pictorially, we place n dots in place p. (See Fig. 1.) The dots in the places represent the initial marking. A (generalized) Petri net may have parallel arcs, that is, arcs with multiplicity greater than one. However, for brevity we will deal only with Petri nets with arc multiplicity equal to one in this article, unless otherwise stated.

A transition t of a Petri net is said to be *firable* or *enabled* if each input place of t is marked with at least one token. For example, in Fig. 2, transition 1 is firable whereas transition 3 is not firable. A *firing* of a (firable) transition removes one token from each input place and adds one token to each output place. In Fig. 2, we illustrate markings which result from successively firing transitions 1, 2, 3, 4, and 5.

In modeling using Petri nets, we regard the places as conditions and the transitions as events. A token in a place (condition) indicates that that condition is met; hence, when all input places for a particular transition (event) have tokens, then all conditions are met for that event and that event can take place. The firing of a transition corresponds to the occurrence of that event.

For example, consider the predictor-corrector computation scheme [**30**] for solving the ordinary differential equation $\dot{x} = f(x, t)$

$$x_{n+1}^p = x_{n-1}^c + 2hf(x_n^p, t_n) \tag{1}$$

$$x_n^c = x_{n-1}^c + \frac{h}{2}\left[f(x_n^p, t_n) + f(x_{n-1}^c, t_{n-1})\right]. \tag{2}$$

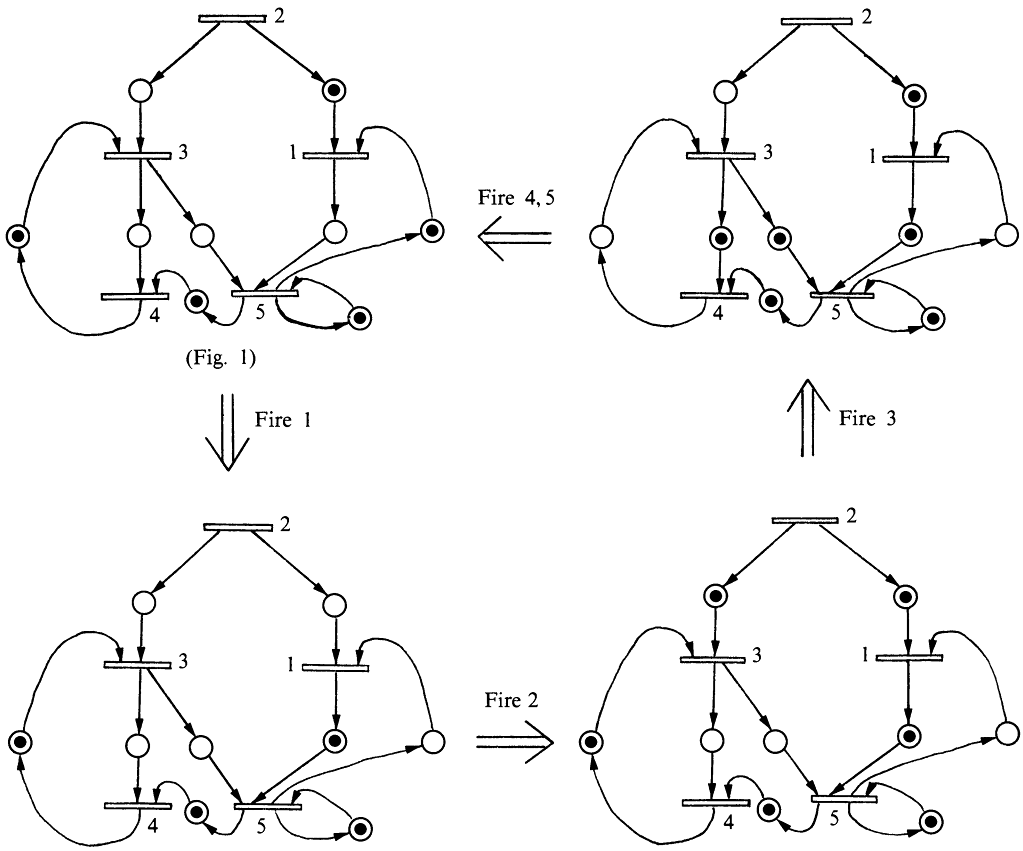


FIG. 2. A sequence of markings for Fig. 1.

This computation can be modeled by the Petri net of Fig. 1. The presence of tokens in the places t_0, x_0^c, and x_1^p means that the conditions, "the values of t_0, x_0^c, and x_1^p have been assigned," are met. In particular, transition 1 can be fired. The firing of this transition, shown in Fig. 2, models the calculation of $f(x_0^c, t_0)$. At this point we cannot fire transition 3 since input place P_0 does not have a token—a condition is not met. To obtain a token at place P_0 we must fire transition 2 which simulates the event, "increment t_n." Now all input places to transition 3 have tokens so that transition 3 is enabled. Firing transition 3 simulates the event "compute $f(x_n^p, t_n)$." Fig. 2 illustrates the firing sequence corresponding to one iteration of the algorithm. This net models the inherent concurrency of the algorithm—transitions 4 and 5 can be fired in either order or concurrently.

Petri nets originated in C. A. Petri's dissertation [**46**]. Petri adopted the viewpoint that the basic phenomena of communication are representable by purely combinatorial means. Thus, he proposed the construction of a net with more practical applicability in the design and programming of information processing machines than does the theory of abstract automata. In the mid 1960's, Petri's work was introduced in the United States and has since been developed, extended, and applied to many areas in computer science.

The early work in the United States was carried out principally by a research group at Applied Data Research (and later at Massachusetts Computer Associates) led by Holt [**17**] and by the Computation Structures Group at M.I.T. [**9**], [**13**], [**41**], [**47**]. At about the same time, Karp and Miller [**22**], [**23**] independently introduced vector addition systems, a parallel computation model closely related to Petri nets. Petri and his colleagues have continued their research and currently more research activities are taking place in Europe than in the United States [**5**], [**52**].

II. Applications. The role of Petri nets in modeling was described in the previous section. Petri nets and other related graph models have been used to investigate a variety of systems such as computer operating systems [39], [40], computer software [9], [25], [26], computer hardware [21], [42], legal systems [28], formal language theory [8], [16], [43], communication protocols [29], PERT charts [44], chemical systems [44], interrelationships of mathematical structures [12], office information systems [11], performance evaluation [49] and fault-tolerant systems [50], industrial process control [3], distributed database systems [51], and the formal verification of parallel programs [24]. In this section we discuss some of these models.

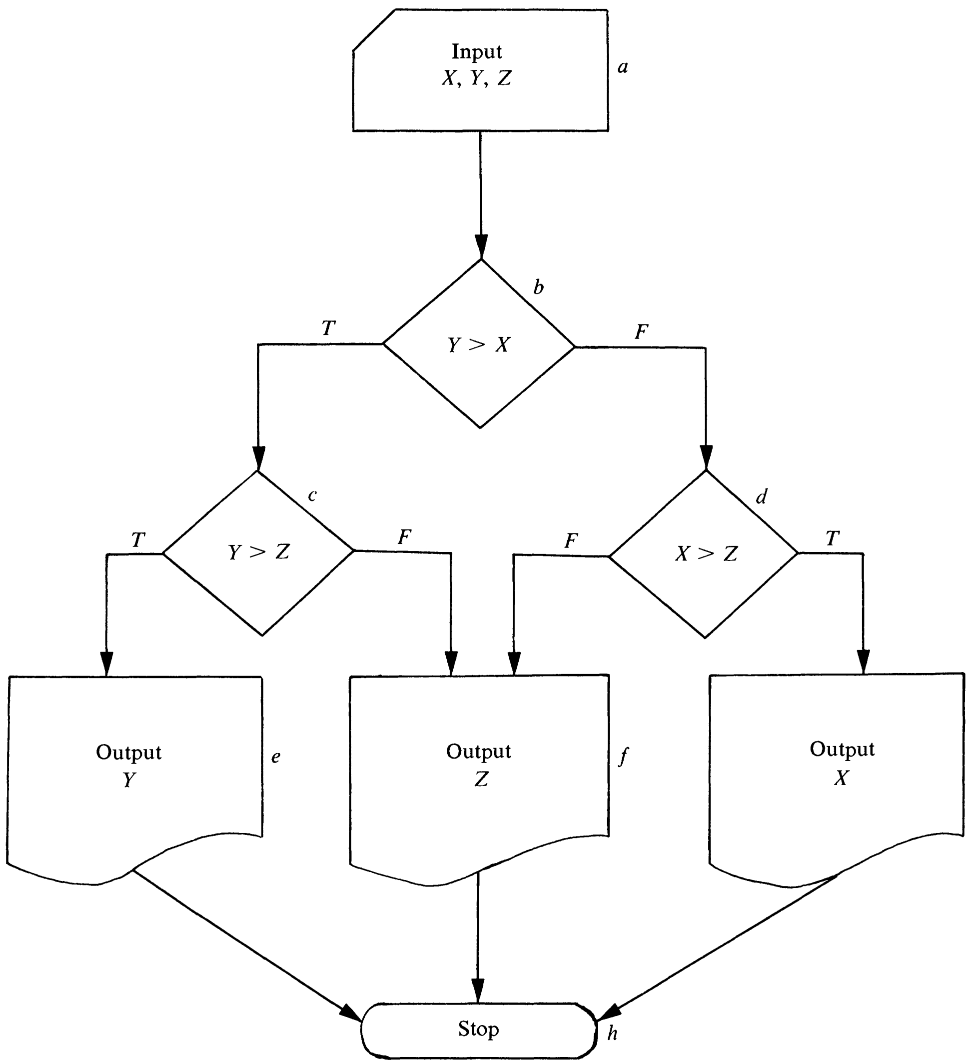


FIG. 3. Flowchart.

In the area of computer software, it is easy to convert a flowchart of a computer program to a Petri net. (See Figs. 3 and 4.) The boxes (respectively arcs) of the flowchart correspond to transitions (respectively places) in the Petri net.

The computer program

$$A = 1, \quad B = 2, \quad C = A{**}2, \quad D = A + B, \quad E = C{*}D$$

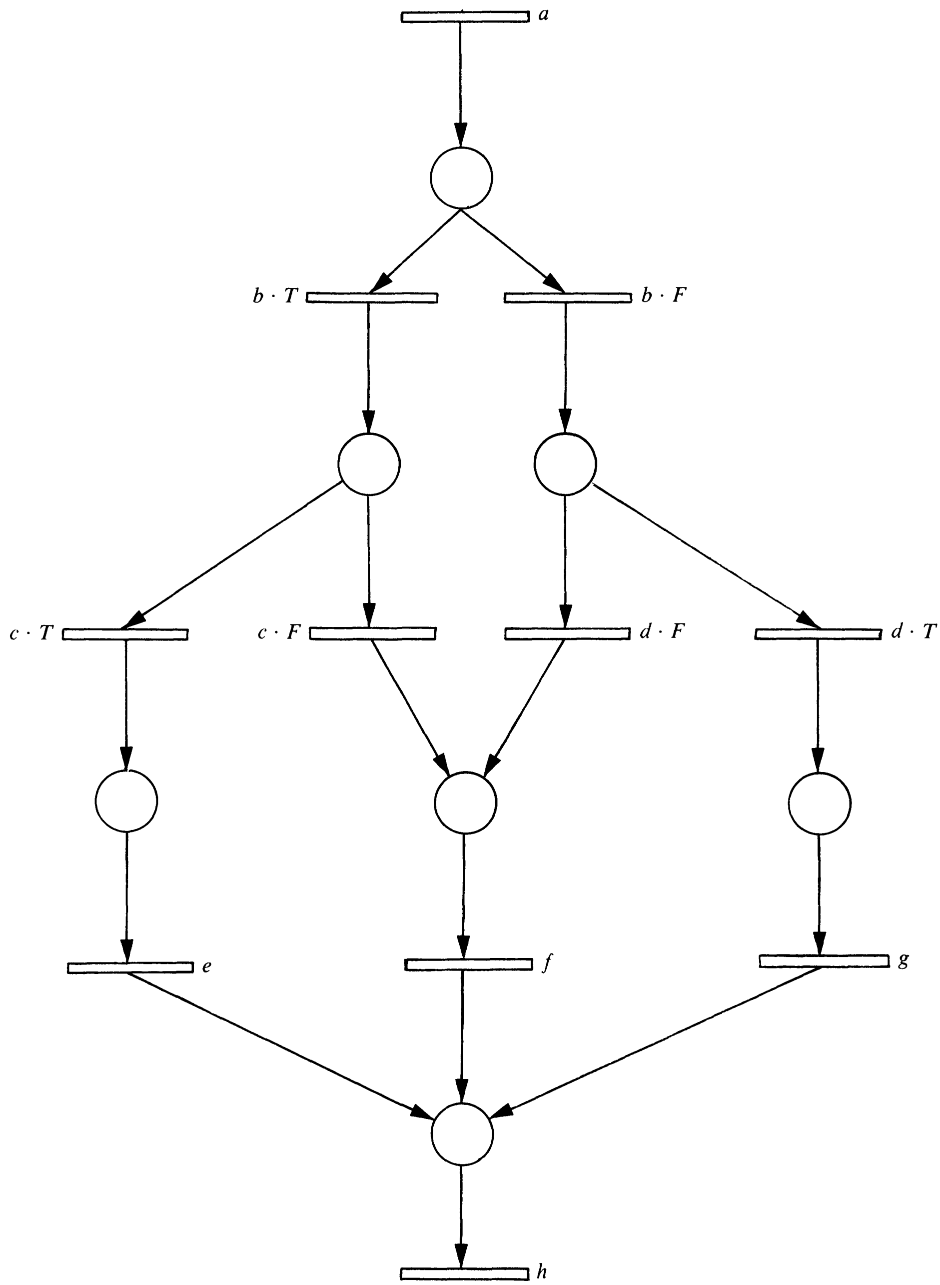


FIG. 4. Flowchart of Fig. 3 as a Petri net.

can be modeled by the Petri net of Fig. 5. This model demonstrates the fact that the statements $C = A{*}{*}2$ and $D = A + B$, for example, can execute in either order or simultaneously. Such models are becoming increasingly important because the decline of hardware costs makes feasible the introduction of multiple processors making possible concurrent processing.

Models similar to that of Fig. 5 have proved useful in compiler design. Having obtained a Petri net model of the instructions to be executed which maximizes parallelism, we can consider all possible execution sequences, together with execution times and/or costs of execution for a

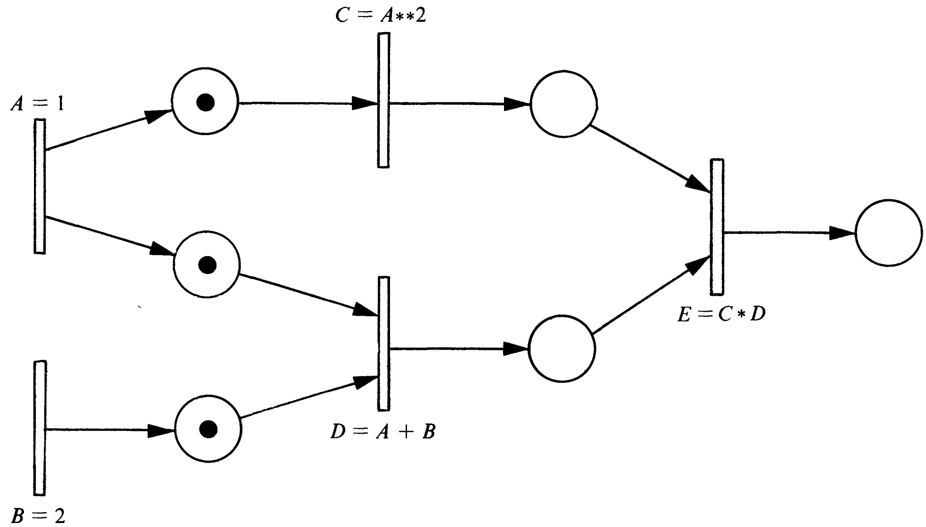


FIG. 5. Petri net model of a computer program.

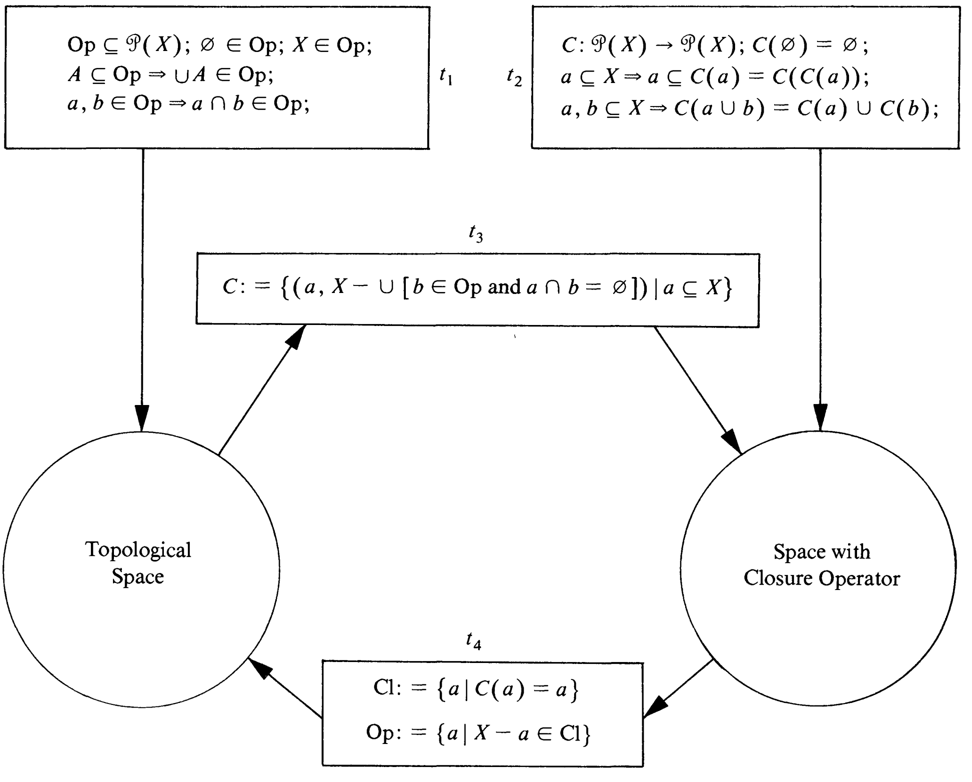


FIG. 6. Petri net representation of some mathematical structures.

particular computer. This allows us to compile optimally. Such an analysis was carried out in the design of a FORTRAN compiler for the CDC 6600 computer [**48**].

Finally, in Fig. 6, we show a Petri representation of some concepts in topology. In this example, adapted from Genrich [**12**], the transitions are represented by rectangular boxes. In

Genrich's theory, a token has structure. For example, we could obtain a token in the place labeled "Topological Space" by firing transition t_1. This token would be assumed to have the structure specified by t_1, namely, that of a topological space. If we next fire t_3, we will produce a token in the place labeled "Space with Closure Operator." We could also obtain a space with a closure operator by firing transition t_2. Thus, a Petri net is used in a formal way to show the interrelationships of mathematical structures.

III. Properties of Petri Nets. Among many other properties, three are characteristic of Petri nets and have been most frequently studied: reachability, liveness, and safeness (or boundedness). A marking M_n is said to be *reachable* from a marking M_0 if there exists a firing sequence which transforms M_0 to M_n. A marking M_0 is said to be *live* for a Petri net if, no matter what marking has been reached from M_0, it is possible ultimately to fire any transition of the net by progressing through some further firing sequence. Thus, a Petri net with a live marking guarantees a deadlock-free operation, regardless of the firing sequence chosen. The markings shown in Figs. 1 and 2 are live. An example of a Petri net with a nonlive marking is given in Fig. 7(a). Here, no transition can be fired after transition t_1, as indicated in Fig. 7(b). (However, the marking shown in Fig. 7(a) may be said to exhibit a weaker form of liveness in the sense that each transition can be fired once if the firing sequence is chosen as t_2, t_4, t_5, t_1, t_3.) A marking M_0 is said to be *bounded* if there exists an integer m such that each place of the net has at most m tokens for every marking reachable from M_0. Specifically, M_0 is said to be *safe* if $m = 1$. The concept of boundedness is related to the capacity bound of memory devices. For example, if registers which can store one word at a time are used as places, then a safe marking will not cause overflow in these registers, regardless of the firing sequence chosen. The marking shown in Fig. 1 is live but not safe, and the markings of Fig. 7 are safe but not live. An example of a Petri net with a live and safe marking M_0 is given in Fig. 8. Here we also show all markings reachable from M_0. These are also live and safe.

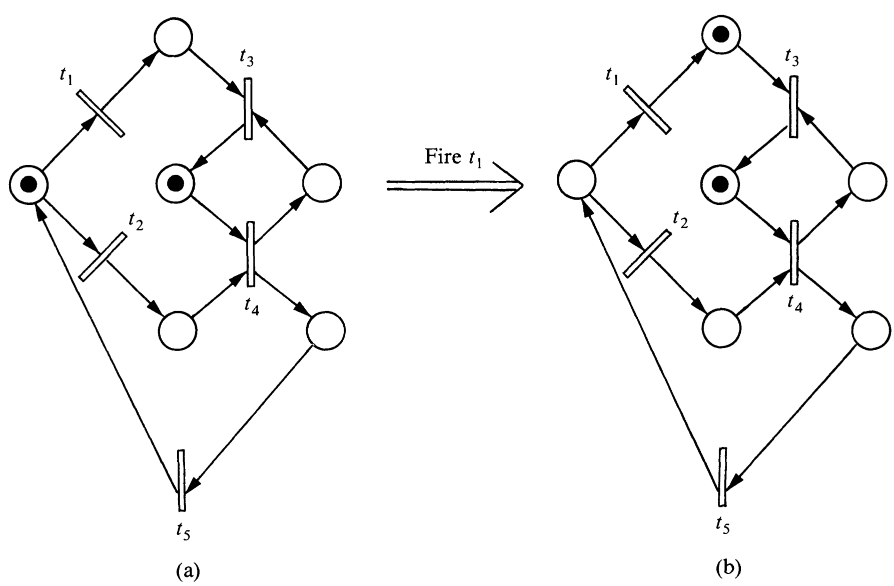


FIG. 7. A Petri net with a non-live marking.

There has been extensive work reported on the formal properties of Petri nets as is shown in a recent survey [**19**]. However, most of the results to date are concerned with subclasses of Petri nets (such as marked graphs and freechoice nets) or partial solutions. While the reachability question for Petri nets (or equivalently, foı vector addition systems [**22**], [**38**]), has recently been reported decidable [**18**], [**27**], there are also undecidable questions concerning Petri nets. For example, it has

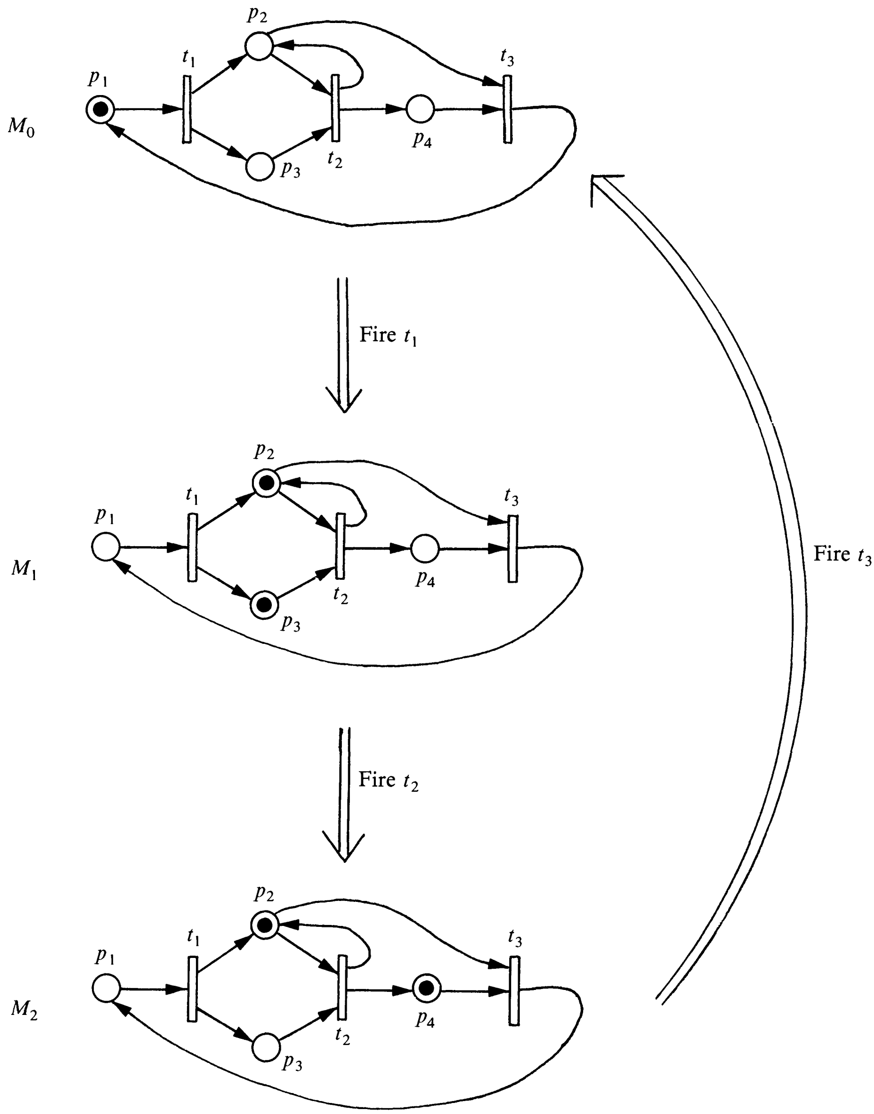


FIG. 8. A live and safe Petri net.

been shown [4], [14] that the question of whether two Petri nets with equal numbers of places have the same reachability set (the set of all markings reachable from an initial marking) is undecidable.

IV. Matrix Approach. Let p and t denote the number of places and transitions in a Petri net, respectively. A *marking* or *state vector* M_k is a $p \times 1$ column vector of nonnegative integers. The jth entry of this vector denotes the number of tokens on place j immediately prior to the kth firing. Specifically, M_0 denotes the initial marking or state. The kth *firing* or *control vector* v_k is a $t \times 1$ column vector of 1's and 0's. Its ith entry is 1 if transition i is to be fired at the kth firing opportunity and 0 otherwise. Let $A^- = [a_{ij}^-]$ $(A^+ = [a_{ij}^+])$ be a $t \times p$ matrix having a_{ij}^- $(a_{ij}^+) = 1$ if place j is an input (output) place for transition i and 0 otherwise. Then, transition i is firable at marking M_k if each entry of M_k is greater than or equal to the corresponding entry of the ith row of A^-. Further, we see that the ith row of the matrix A, which is defined as $A^+ - A^-$

represents the token change in each of the p places when transition i fires once. In other words, the marking M_{k+1} resulting from marking M_k by the kth firing v_k may be given in terms of the following matrix (state) equation [**34**]:

$$M_{k+1} = M_k + A^T v_k, \qquad k = 0,1,2,\ldots \tag{3}$$

where T denotes matrix transposition. It should be noted that the control vector v_k cannot be chosen arbitrarily since only enabled transitions may be fired. Also, observe that v_k is constrained by $M_k + A^T v_k \geqslant 0$ for each k. Obviously, matrix A in (3) is of key importance in characterizing both the dynamic behavior and the structure of a Petri net. The usefulness of the state equation (3) is demonstrated in the next section of this article for a subclass of Petri nets and in [**31**], [**34**] for more general nets.

The matrices A^- and A^+ of the net given in Fig. 8 are

$$A^- = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array} \!\!\begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

and

$$A^+ = \begin{array}{c} t_1 \\ t_2 \\ t_3 \end{array} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Equation (3) is illustrated below for the transition to the marking $M_2 = (0\ \ 1\ \ 0\ \ 1)^T$ from $M_1 = (0\ \ 1\ \ 1\ \ 0)^T$, caused by firing t_2.

$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{array}{c} p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \begin{bmatrix} -1 & 0 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

State equation (3) can be extended to generalized (multiple arc) Petri nets. Further, a state equation can be written [**37**] for E-nets, modified Petri nets, which have been used for modeling the CDC 6400 computer under the control of an operating system [**39**].

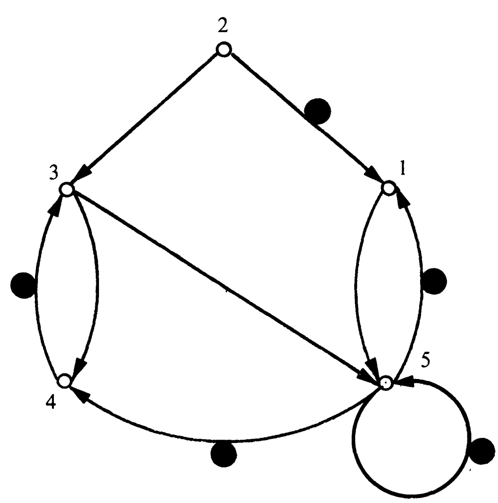


FIG. 9. A marked graph representation of the Petri net of Fig. 1.

V. Marked Graphs. In many applications, as for example in Fig. 1, each place in a Petri net has exactly one incoming arc and exactly one outgoing arc. Such a Petri net can be reduced to a directed graph, where vertices correspond to transitions and arcs to places. This subclass of Petri nets is known as *marked graphs*. As an example, the marked graph representation of the Petri net of Fig. 1 is drawn in Fig. 9. The tokens are shown on the arcs.

For marked graphs, properties such as liveness, safeness, and reachability have been completely characterized.

Liveness and Safeness Conditions for Marked Graphs. The firing of a vertex (transition) in a marked graph removes one token from each of its incoming arcs (input places) and adds one token to each of its outgoing arcs (output places). If a vertex is on a directed circuit, then exactly one of its incoming arcs and exactly one of its outgoing arcs belongs to the directed circuit. If a vertex does not lie on the directed circuit in question, none of the arcs incident to that vertex will belong to the directed circuit. Thus, we have the following token invariance property [**7**].

Property 1. The token count in a directed circuit is invariant under any vertex firing.

By Property 1, if there are no tokens on a directed circuit at an initial marking, then this directed circuit remains token-free, and the vertices on this directed circuit must forever be nonfirable. It can also be shown [**7**] that if a vertex cannot be made firable by any firing sequence, then this vertex must belong to a token-free directed circuit. Therefore, we have:

Property 2. A marking M_0 is live for a directed graph G if and only if M_0 places at least one token in each directed circuit in G.

The following property is a mini-max theorem of marked graphs. (See eq. (4) in [**35**].)

Property 3. The maximum number of tokens that an arc can have in a marked graph G, with an initial marking M_0, is equal to the minimum number of tokens placed by M_0 on a directed circuit containing this arc.

From Property 3 we can conclude that: (1) If an arc belongs to a token-free directed circuit, then this arc will never possess tokens and (2) if an arc belongs to a directed circuit having a token count of one, and does not belong to any token-free directed circuit, then this arc will have at most one token. In fact, it is known [**7**] that:

Property 4. A live marking is safe for a marked graph G if and only if every arc in G belongs to a directed circuit with token count one.

It is interesting to note that live and safe markings are related to minimal (but not minimum) feedback arc sets [**2**], and that the maximum live and safe marking is akin to a maximum-cost minimum-flow problem [**7**].

Fig. 10 shows a marked graph G with a live initial marking M_0. The two markings reachable from M_0 are also illustrated. There are three directed circuits (ac, de, abd) in G. The token counts of these directed circuits are 2 (on ac), 1 (on de), and 1 (on abd). Observe that these counts do not change with any firing (Property 1). The marking M_0 is live since each directed circuit has at least one token (Property 2), but it is not safe because arc c does not belong to a directed circuit with token count one (Property 4). There are two directed circuits, ac and abd, which include arc a. The maximum number of tokens that arc a can have is given by the minimum of two (for ac) and one (for abd), or one, as verified by Fig. 10 (Property 3). A live and safe marking for the same graph G is shown in Fig. 11.

Reachability Theorem for Marked Graphs. To discuss the reachability theorem for marked graphs, we need the *fundamental circuit matrix* B_f of a directed graph G [**10**]. Momentarily, consider G to be an undirected graph. Select a spanning tree T of G; that is, a connected, circuit-free subgraph T of G containing all of the vertices of G. For example, (ab) is a spanning

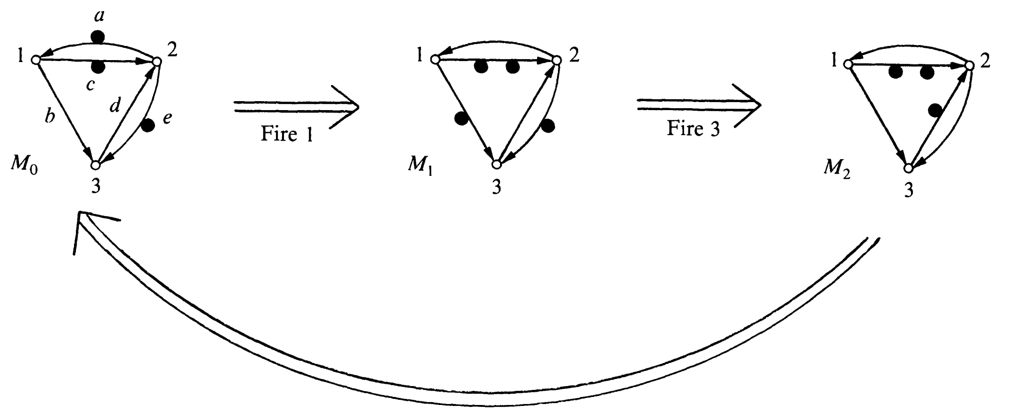


FIG. 10. A live and unsafe marked graph.

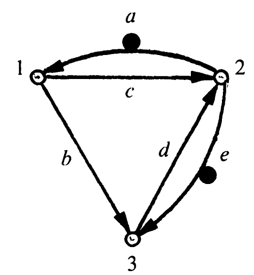


FIG. 11. A live and safe marked graph.

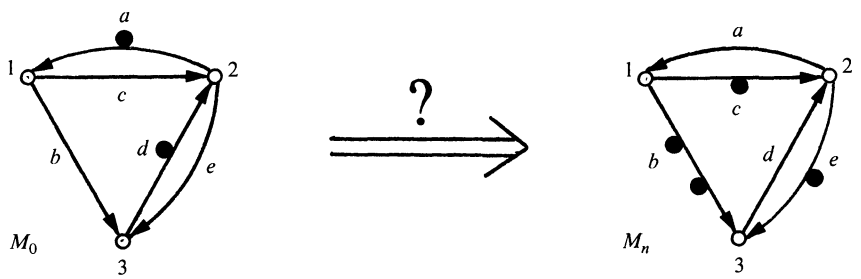


FIG. 12. Application of equation (6).

tree of the graph of Fig. 12. Note that adding any edge not in T to T determines a unique (undirected) circuit, called a *fundamental circuit*.

The rows of the fundamental circuit matrix B_f correspond to the fundamental circuits and the columns of B_f correspond to the edges of G. Each fundamental circuit is assigned an orientation consistent with the orientation of the added arc which determines it. The row entries in B_f corresponding to a fundamental circuit C are defined as follows. If an edge is not in C, the entry is 0. If an edge is in C and is oriented in the same way as C, the entry is 1; if the orientation is reversed, the entry is -1. The fundamental circuit matrix for Fig. 12 is given below.

$$\begin{array}{c} \\ (ac) \\ B_f = (abd) \\ (abe) \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Suppose that there exists a firing sequence $\{v_0, v_1, \ldots v_{n-1}\}$ that transforms an initial marking M_0 to M_n in a marked graph G. Applying (3) for $k = 0, 1, \ldots, n-1$ and summing, we obtain

$$M_n = M_0 + A^T \sum_{k=0}^{n-1} v_k,$$

which may in turn be written as

$$A^T X = \Delta M \tag{4}$$

where $\Delta M = M_n - M_0$ and X, the *firing count vector*, is defined as $\sum_{k=0}^{n-1} v_k$. Note that X is a $t \times 1$ column vector of nonnegative integers. The ith entry of X denotes the number of times that vertex i fires in the firing sequence leading from M_0 to M_n. Now let B_f be the fundamental circuit matrix of the directed graph G. It is well known [**10**] that

$$B_f A^T = 0. \tag{5}$$

From (4) and (5), we have

$$B_f \Delta M = 0, \tag{6}$$

or

$$B_f M_0 = B_f M_n. \tag{7}$$

Equation (7) states that the algebraic sum of tokens on a fundamental circuit placed by M_0 is equal to that placed by M_n, and is a generalization of Property 1. It has been shown [**32**] that this generalized token invariance expressed by (6) is not only necessary but also sufficient for a live marking M_0 to reach another marking M_n. In other words, we have:

Property 5. In a marked graph, a marking M_n is reachable from a live marking M_0 if and only if (6) holds.

Condition (6) for Fig. 12, when (ab) is taken as our tree, becomes

$$B_f \Delta M = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Thus, M_n is reachable from M_0. A firing count vector X can be found by solving (4) for Fig. 12:

$$\begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

Since the rank of the coefficient matrix is two, we may set $x_3 = 0$ and solve for x_1 and x_2 obtaining $X = (2\ 1\ 0)^T$. Therefore, M_0 reaches M_n by firing vertex 1 twice and vertex 2 once (in the sequence 1-2-1 as the reader may verify with reference to Fig. 12).

Property 5 can be extended to the case of a nonlive initial marking M_0 if one imposes the additional condition that the vertices which are to fire should not lie on a token-free directed circuit [**32**].

Another application of (6) is the following. Since ΔM is zero if and only if all the entries in X are the same, we have:

Property 6. A firing sequence leads back to the initial marking ($M_n = M_0$) if and only if it fires every vertex an equal number of times.

Fig. 10 provides an illustration of Property 6.

Controllability. The controllability problem is: Given any initial marking M_0 and any marking M_n, find a condition that there exists a firing sequence leading from M_0 to M_n. Since ΔM is a $p \times 1$ matrix, if (4) can be solved for X for any choice of ΔM, we must have

$$\operatorname{Rank} A = p. \tag{8}$$

Although condition (8) is only necessary for the existence of a firing sequence leading from any M_0 to any M_n within unrestricted Petri nets [**34**], it is necessary and sufficient in the case of marked graphs as we will show.

Without loss of generality, we may assume that our marked graph G is connected. In this case, A is simply the incidence matrix of G and it is known [**10**] that for the incidence matrix A of a connected graph with t vertices

$$\operatorname{Rank} A = t - 1. \tag{9}$$

From (8) and (9), we have

$$p = t - 1 \tag{10}$$

which implies that G contains no circuits and is a tree. For a marked graph, the underlying graph of which is a tree, any initial marking M_0 is live since there is no directed circuit (Property 2) and (6) holds for any ΔM. Therefore, we have [**32**]:

Property 7. There exists a firing sequence leading from any initial marking to any other marking in a marked graph G if and only if the underlying graph of G is a tree.

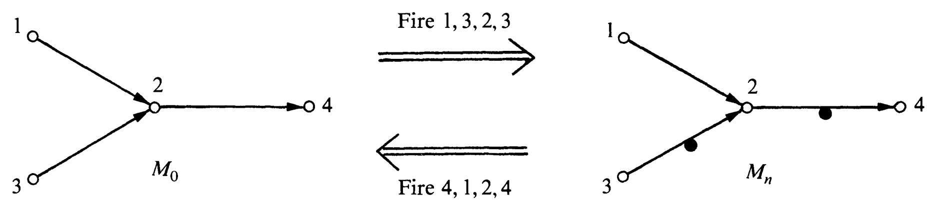


FIG. 13. A marked tree graph.

In the marked (tree) graph shown in Fig. 13, the firing sequence 1-3-2-3 transforms M_0 to M_n, while the sequence 4-1-2-4 drives M_n back to M_0. Note that in a marked tree graph, there is always at least one source vertex (vertices 1 and 3 in Fig. 13) which can create tokens, and there will always be at least one sink vertex (vertex 4 in Fig. 13) which can remove tokens from the graph.

Define a relation $\sim$ on the set of live markings of a marked graph G to be $M_0 \sim M_n$ if M_n is reachable from M_0. Then $\sim$ is an equivalence relation and thus partitions the set of live markings into equivalence classes. For example, three different equivalence classes of markings for the same directed graph are shown in Figs. 10, 11, and 12. Let $\rho(G)$ be the number of equivalence classes of live and safe markings for a directed graph G. Finding $\rho(G)$ in the general case is an unsolved problem of marked graphs; however, several transformations of marked graphs have been introduced which are useful in finding $\rho(G)$ and for analyzing other properties [**20**], [**36**]. These transformations have also been used in synthesizing marked graphs having prescribed properties [**35**].

References

1. T. Agerwala, Putting Petri nets to work, Computer, vol. 12 (December 1979) 85–94.

2. A. T. Amin and T. Murata, A characterization of live and safe markings of a directed graph, Proc. of the 1976 Conf. on Info. Sci. and Syst., Johns Hopkins University (March 1976) 295–299.

3. C. Andre, M. Diaz, C. Girault, and J. Sifakis, Survey of French research and applications based on Petri nets, pp. 321–345, in W. Brauer (Editor), Net Theory and Applications, Lecture Notes in Computer Science, vol. 84, Springer-Verlag, New York, 1980.

4. T. Araki and T. Kasami, Undecidable problems for Petri nets, Systems-Comput.-Controls, vol. 7, no. 1 (January-February 1976) 20–28.

5. W. Brauer (Editor), Net Theory and Applications, Lecture Notes in Computer Science, vol. 84, Springer-Verlag, New York, 1980.

6. C. T. Chen, Introduction to Linear System Theory, Holt, Rinehart and Winston, New York, 1970.

7. F. Commoner, A. W. Holt, S. Even, and A. Pnueli, Marked directed graphs, J. Comput. System Sci., vol. 5 (October 1971) 511–523.

8. S. Crespi-Reghizzi and D. Mandrioli, Some algebraic properties of Petri nets, Alta Frequenza, vol. 45, no. 2 (February 1976) 130–137.

9. J. B. Dennis, Concurrency in software systems, in Advanced Course in Software Engineering, F. L. Bauer (Editor), Springer-Verlag, Berlin, 1973, pp. 111–127.

10. N. Deo, Graph Theory with Applications to Engineering and Computer Science, Prentice-Hall, Englewood Cliffs, N. J., 1974.

11. C. A. Ellis and G. J. Nutt, Office information systems and computer science, ACM Computing Surveys, vol. 12 (1980) 27–60.

12. H. J. Genrich, The Petri net representation of mathematical knowledge, St. Augustin, West Germany, GMD-ISF Report ISF-76-05, 1976.

13. M. Hack, Analysis of production schemata by Petri nets, M.I.T., Project MAC, TR-94, 1972.

14. M. Hack, Decidability questions for Petri nets, Technical Report TR-161, M.I.T. Lab. for Comp. Sci., June 1976.

15. M. Hack, Decision problems for Petri nets and vector addition problems, Computation Structures Group Memo 95, Project MAC, M.I.T., Cambridge, Mass., March 1974; also Technical Memo 59, Project MAC, M.I.T., March 1975.

16. M. Hack, Petri net languages, Rep. TR 159, M.I.T., Lab. for Comp. Sci., March 1976.

17. A. W. Holt, et al., Final report of the information system theory project, Tech. Rept. RADC-TR-68-305, Rome Air Development Center, 1968.

18. S. Hudak, On the reachability problem for vector addition systems, presented at 2nd European Workshop on Applications and Theory of Petri Nets, Bad Honnef, Germany, September 28–30, 1981.

19. M. Jantzen and R. Valk, Formal properties of place/transition nets, pp. 165–212 in W. Brauer (Editor), Net Theory and Applications, Lecture Notes in Computer Science, vol. 84, Springer-Verlag, New York, 1980.

20. R. Johnsonbaugh and T. Murata, Additional methods for reduction and expansion of marked graphs, IEEE Trans. Circuits and Systems, vol. CAS-28, no. 10 (October 1981) 1009–1014.

21. J. R. Jump, Asynchronous control arrays, IEEE Trans. Comput., vol. C-23, no. 10 (October 1974) 1020–1029.

22. R. M. Karp and R. E. Miller, Parallel program schemata, J. of Comp. and Syst. Sci., vol. 3, no. 2 (May 1969) 147–195.

23. R. M. Karp and R. E. Miller, Properties of a model for parallel computations: determinacy, termination, queueing, SIAM J. Appl. Math., vol. 14 (November 1966) 1390–1411.

24. R. M. Keller, Formal verification of parallel programs, Comm. ACM, 19 (1976) 371–384.

25. P. E. Lauer and R. H. Campbell, Formal semantics of a class of high-level primitives for coordinating concurrent processes, Acta Inform., vol. 5, no. 4 (1975) 297–332.

26. K. Lautenbach and H. A. Schmid, Use of Petri nets for proving correctness of concurrent process systems, Inform. Process. '74, North-Holland, Amsterdam, 1974, pp. 187–191.

27. E. W. Mayr, An algorithm for the general Petri net reachability problem (preliminary version), Proc. of ACM Symposium on Theory of Computing, University of Wisconsin-Milwaukee, May 1981.

28. J. A. Meldman, A Petri net representation of civil procedure, IDEA: The Journal of Law and Technology, vol. 19, no. 2 (1978) 123–148.

29. P. M. Merlin, A methodology for the design and implementation of communication protocols, IEEE Trans. Comm., vol. COM-24 (June 1976) 614–621.

30. R. E. Miller, Some relationships between various models of parallelism and synchronization, IBM Res. Report RC-5074 (October 1974).

31. G. Memmi and G. Roucairol, Linear algebra in net theory, in Net Theory and Applications, Lecture Notes in Comp. Sci., vol. 84, Springer-Verlag, New York, 1980, pp. 213–224.

32. T. Murata, Circuit theoretic analysis and synthesis of marked graphs, IEEE Trans. on Circuits and Syst., vol. CAS-24, no. 7 (July 1977) 400–405.

33. T. Murata, Petri nets, marked graphs, and circuit-system theory—a recent CAS application, Circuits and Systems, vol. 11, no. 3 (June 1977) 2–12.

34. T. Murata, State equation, controllability, and maximal matchings of Petri nets, IEEE Trans. Auto. Control, vol. AC-22, no. 3 (June 1977) 412–416.

35. T. Murata, Synthesis of decision-free concurrent systems for prescribed resources and performance, IEEE Trans. on Software Engrg., vol. SE-6, no. 6 (November 1980) 525–530.

36. T. Murata and J. Y. Koh, Reduction and expansion of live and safe marked graphs, IEEE Trans. on Circuits and Systems, vol. CAS-27 (January 1980) 68–70.

37. T. Murata and T. Shah, On liveness, deadlock, and reachability of *E*-nets, Procs. of the 1976 Allerton Conf. on Circuit and System Theory, October, 1976.

38. B. O. Nash, Reachability problems in vector addition systems, this MONTHLY, vol. 80 (1973) 292–295.

39. J. D. Noe and G. J. Nutt, Macro *E*-nets for representation of parallel systems, IEEE Trans. on Comp., vol. C-22, no. 8 (August 1973) 718–727.

40. G. J. Nutt, Evaluation nets for computer system performance analysis, 1972 Fall Joint Comp. Conf., AFIPS Conf. Proc., vol. 41, Montvale, New Jersey: AFIPS Press, 1972, pp. 279–286.

41. S. S. Patil, Coordination of asynchronous events, M.I.T., Project MAC, Tech. Report 72, 1970.

42. S. S. Patil, Micro-control for parallel asynchronous computers, Comp. Struct. Group Memo 120, Project MAC, M.I.T., March 1975.

43. J. L. Peterson, Computation sequence sets, J. of Comp. and Syst. Sci., vol. 13, no. 1 (August 1976) 1–24.

44. J. L. Peterson, Petri Net Theory and the Modeling of Systems, Prentice-Hall, Englewood Cliffs, NJ, 1981.

45. J. L. Peterson, Petri nets, Computing Surveys, vol. 9 (September 1977) 223–252.

46. C. A. Petri, Kommunikation mit Automaten (in German), University of Bonn, 1962; translation by C. F. Green, Jr. Communication with Automata, Supplement to Tech. Doc. Rep. #1, Rome Air Development Center, Contract #AF30(602)-3324, 1965.

47. C. Ramchandani, Analysis of asynchronous concurrent systems by timed Petri nets, M.I.T., Project MAC, Tech. Report 120, 1974.

48. R. M. Shapiro and H. Saint, A new approach to optimization of sequencing decisions, Ann. Rev. of Automatic Programming, vol. 6, no. 5 (1970) 257–288.

49. J. Sifakis, Use of Petri nets for performance evaluation, in Modelling and Performance Evaluation of Computer Systems, edited by Beilner and Gelenbe, North-Holland, Amsterdam, 1977, pp. 75–93.

50. J. Sifakis, Realization of fault-tolerant systems by coding Petri nets, J. Design Automat. Fault-Tolerant Comput., vol. 3, no. 2, 1979.

51. K. Voss, Using predicate/transition-nets to model and analyze distributed database systems, IEEE Trans. on Software Engineering, vol. SE-6, no. 6 (November 1980) 539–544.

52. C. Girault and W. Reisig (Editors), Application and Theory of Petri Nets, Informatik-Fachberichte, vol. 52, Springer-Verlag, New York, 1982.

MISCELLANEA

82. In the eyes of a modern mathematician, individual "pretty theorems" have even less intrinsic value than the discovery of a new "pretty flower" has for a scientific botanist. However, amateurs find the principal appeal of the respective sciences in such things.

If you will look at mathematical journals that have relaxed their standards, you will see what rank weeds spring up in the absence of the shears of the critical gardener and of the question, *quid usui*?

I need hardly say that the utility of theorems does not need to be sought *outside* mathematics. It is true that both the astronomy of the last century and today's physics, with its mathematical foundation, have raised and continue to raise many problems that have influenced the development of mathematics and have led to the creation of splendid theories, which would never have been discovered in the course of purely mathematical activity. Nevertheless, it is certain that the development of the main branches of our science has, since the beginning, come about from intrinsic necessity.

—Hermann Hankel, *Die Entwickelung der Mathematik in den letzten Jahrhunderten*, Tübingen, 1869, p. 19.

How many mathematicians do you know whose name is attached to four major theorems? See p. 588.

is already proved for the dimension $n-1$. I could almost stop here!

CLAIM. *If U is a neighborhood of 0 in $\mathbb{C}^n$, F: $U \to \mathbb{C}^n$ is holomorphic and one-to-one, and $(\partial f_i/\partial z_j)(0) \neq 0$ for some i and j, then* (JF) (0) $\neq 0$.

Proof of the Claim. Without loss of generality we can assume that $F(0)=0$ and $(\partial f_1/\partial z_1)(0) = a \neq 0$. Define $H(z) = (f_1(z), z_2, \ldots, z_n)$. Since (JH)(0) $= a \neq 0$, H is a biholomorphic map of a neighborhood of 0 onto a neighborhood V of 0. Therefore

$$G(w) = F(H^{-1}(w))$$

defines a holomorphic one-to-one map $G = (g_1, \ldots, g_n)$ of V into $\mathbb{C}^n$. For $w = (w_1, \ldots, w_n)$ in V, we see that

$$g_1(w) = f_1(H^{-1}(w)) = w_1$$

and (JF)(0) $= a$(JG)(0). Setting $w' = (w_2, \ldots, w_n)$, define

$$\tilde{G}(w') = (g_2(0,w'), \ldots, g_n(0,w')).$$

Then $\tilde{G}$ is holomorphic and one-to-one in some neighborhood of $0'$ in $\mathbb{C}^{n-1}$, so that $(\mathrm{J}\tilde{\mathrm{G}})(0') \neq 0$ by induction hypothesis. But $(\mathrm{J}\tilde{\mathrm{G}})(0') = $ (JG)(0). Hence (JF)(0) $\neq 0$, as desired.

Proof of the Proposition. If not empty, the set defined by (JF) $= 0$ is an analytic variety of complex codimension 1. On this variety, by the claim, the partial derivatives of all the functions f_i are identically 0. Therefore F is locally constant on this variety (this fact is really elementary, at least at the regular points of the variety), which of course violates the hypothesis that F is one-to-one.

REMARK. We can even avoid any consideration of analytic varieties, giving another proof.

Second proof of the Proposition. We can assume that Ω is bounded, $0 \in \Omega$, and $F(0)=0$. We want to prove that (JF)(0) $\neq 0$. Let g be the function defined by $g(z) = \Sigma_{i=1}^n |f_i(z)|^2$ and if necessary replace Ω by a slightly smaller set so that g is bounded away from 0 near $\partial\Omega$. Choose $\epsilon > 0$ small enough so that the closure of $\{z \in \Omega: g(z) < \epsilon\}$ is contained in Ω. By Sard's Theorem, there exists $\alpha \in (0, \epsilon)$ such that grad $g \neq 0$ on $g^{-1}(\alpha)$. Let $\mathcal{O} = \{z \in \Omega: g(z) < \alpha\}$. The claim implies that JF $\neq 0$ on $\partial\mathcal{O}$. Thus JF $\neq 0$ on $\mathcal{O}$, because for $n > 1$ the zeroes of a holomorphic function propagate out to the boundary (a basic and easy fact). In particular, (JF)(0) $\neq 0$, as desired.

References

1. S. Bochner, W. Martin, Functions of Several Complex Variables, Princeton Univ. Press, Princeton, N.J., 1948.
2.. B. Fuks, Introduction to the theory of analytic functions in several complex variables, Translations of Math. Monographs, v. 8, Amer. Math. Soc., 1963.
3. R. Narasimhan, Several complex variables, Chicago Lect. in Math., 1971.
4. W. Rudin, Function theory in the unit ball of $\mathbb{C}^n$, Grund. der Math. W. 241, Springer Verlag, 1980.
5. (Added in proof) S. Krantz, Function Theory of Several Complex Variables, Wiley, 1982, p. 397.

ANSWER TO "PHOTO" ON PAGE 567

M. H. Stone's name is visible in Stone's theorem (about unitary groups), in the Stone representation theorem (about Boolean algebras), in the Stone-Čech compactification, and in the Stone-Weierstrass theorem. The photo was taken in March, 1982.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

CAN ALL TILES OF A TILING HAVE FIVE-FOLD SYMMETRY?*

LUDWIG DANZER
Department of Mathematics, Universität Dortmund, 4600 Dortmund 50, West Germany

BRANKO GRÜNBAUM*
Department of Mathematics, University of Washington GN-50, Seattle, WA 98195

G. C. SHEPHARD
University of East Anglia, Norwich NR4 7TJ, England

It is well known that the plane can be tiled by equilateral triangles, by squares, or by regular hexagons, but that there exists no such tiling by regular pentagons. More generally it is not hard to see that the plane cannot be tiled by congruent figures each with 5-fold symmetry, that is, admitting as symmetries rotations about a point through angles of 72°, 144°, 216° and 288°. However, if one does not insist that the tiles are congruent, and one allows a variety of different shapes and sizes, then the question of tiling by tiles with 5-fold symmetry becomes much more interesting and leads to several open problems.

We recall that one of the basic results in geometric crystallography is the *crystallographic restriction*, which asserts that in the Euclidean plane no figure with a discrete symmetry group can have more than one center of five-fold rotational symmetry. Similar assertions can be made about k-fold rotations with $k \geqslant 7$. These facts lead to the well-known result that if a planar figure admits translational symmetries but does not admit arbitrarily short translations as symmetries, then the only possible rotational symmetries are k-fold rotations with $k = 2$, 3, 4 or 6 (see, for example, Buerger [**3**, p. 33], Coxeter [**5**, Sect. 4.5], Fejes Tóth [**9**, Sect. I.1.4]).

However, the crystallographic restriction does not prove the impossibility of a tiling in which each individual tile has 5-fold symmetry. Fruitless efforts by mathematicians to construct such tilings go back at least to Kepler [**11**]. Fig. 1 shows some of Kepler's attempts; he notes that in each case a tiling can be obtained only at the price of introducing "monsters" (such as the "fused decagons" in Fig. 1a) which do not have 5-fold symmetry. Explanations of Kepler's tilings and variants of them have been discussed by several authors (Caspar [**4**, p. 374], Bindel [**1**], Eberhart [**7**], Grünbaum & Shephard [**10**, Sect. 2.5]). Even earlier attempts at constructing tilings with tiles that have 5-fold symmetry are discernible in Islamic art (see, for example, Bourgoin [**2**], Critchlow [**6**], Wade [**14**], El-Said & Parman [**8**]).

It is therefore somewhat surprising to find that such tilings can easily be constructed by a simple inductive procedure. In the example indicated in Fig. 2, the tiles used are congruent either to the pentagon A shown in Fig. 2a, or to the nonconvex 20-sided polygon B shown in Fig. 2b, or to a polygon 3^nB similar to B in ratio 3^n, where $n = 1, 2, 3, \ldots$. The construction of the tiling, indicated in Fig. 2c, is based on the fact that five copies of A yield with B a pentagon 3A similar to A in ratio 3; five of these and 3B yield a pentagon 3^2A, etc.

*Research supported by the National Science Foundation Grant MCS77-01629 A01 and a Guggenheim Fellowship.

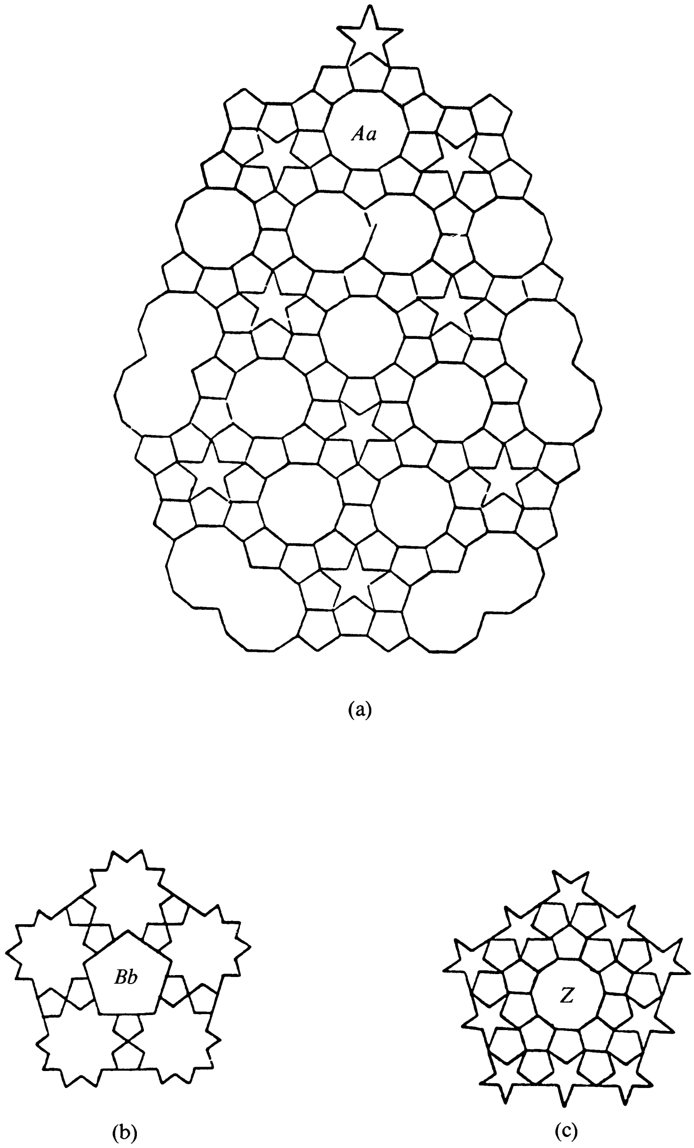


FIG. 1. Some of the patches devised by Kepler in 1619 in attempts to tile the plane by tiles with 5-fold symmetry.

The tiles in this example are not uniformly bounded in diameter. It seems that the answer to the question changes if the diameters of the tiles cannot grow without bounds. To make this problem precise from now on we restrict attention to tilings in which the tiles are closed topological disks of diameter at most 1, covering the plane without gaps or overlaps of their interiors. We conjecture that *the Euclidean plane admits no tiling in which each tile has five-fold symmetry*. If true, this conjecture is precariously balanced against various indications that seem to lend support to the opposite view. For example, tilings by congruent regular pentagons are possible on the sphere, in the elliptic plane and in the hyperbolic plane. The plane can be tiled by affinely regular pentagons, all mutually similar and of only two sizes (see Fig. 3). Many kinds of equilateral convex pentagons—some very close in shape to regular pentagons—can be used to tile

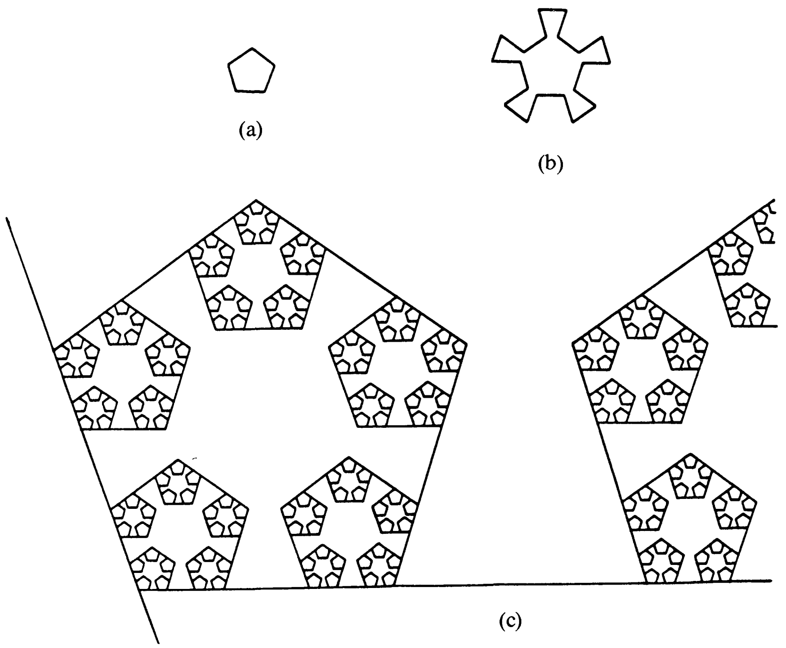


FIG. 2. The construction of a tiling (c) in which each tile is congruent either to the regular pentagon in (a), or to the 20-sided polygon in (b), or to an enlarged version of this polygon.

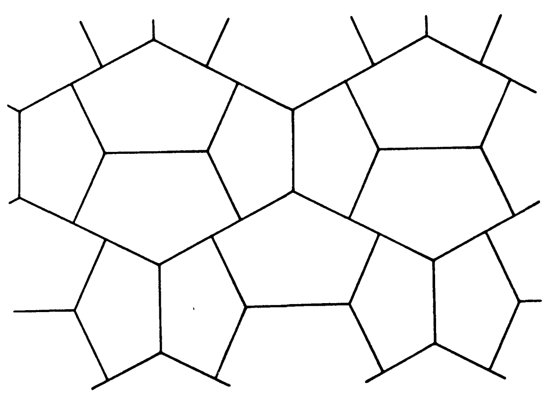


FIG. 3. A tiling of the plane by affinely regular pentagons, all mutually similar and of only two sizes.

the plane by congruent copies (see the surveys in Schattschneider [**12**], [**13**]). Probably even more telling are examples like the one in Fig. 4, which *appears* to show a tiling in which each tile is a regular pentagon. Actually, the pentagons in Fig. 4 form a Cantor-type set, and do not cover the plane; hence they do not form a tiling in the sense considered here.

Large regions of the plane can be covered, in many ways, by *patches* of tiles, in which each tile has 5-fold symmetry or even reflective 5-fold symmetry. In order to compare in a meaningful way the sizes of such patches, various measures can be used. Possibly the most appropriate one is the ratio ρ of the diameter of the largest circular disk covered by the patch to the diameter of the smallest circular disk that can cover each of the tiles in the patch. By adding around Kepler's patch in Fig. 1b ten of the larger pentagons and fifteen of the smaller ones, a patch with $\rho \approx 3.1$ is

[Continued on p. 583.]

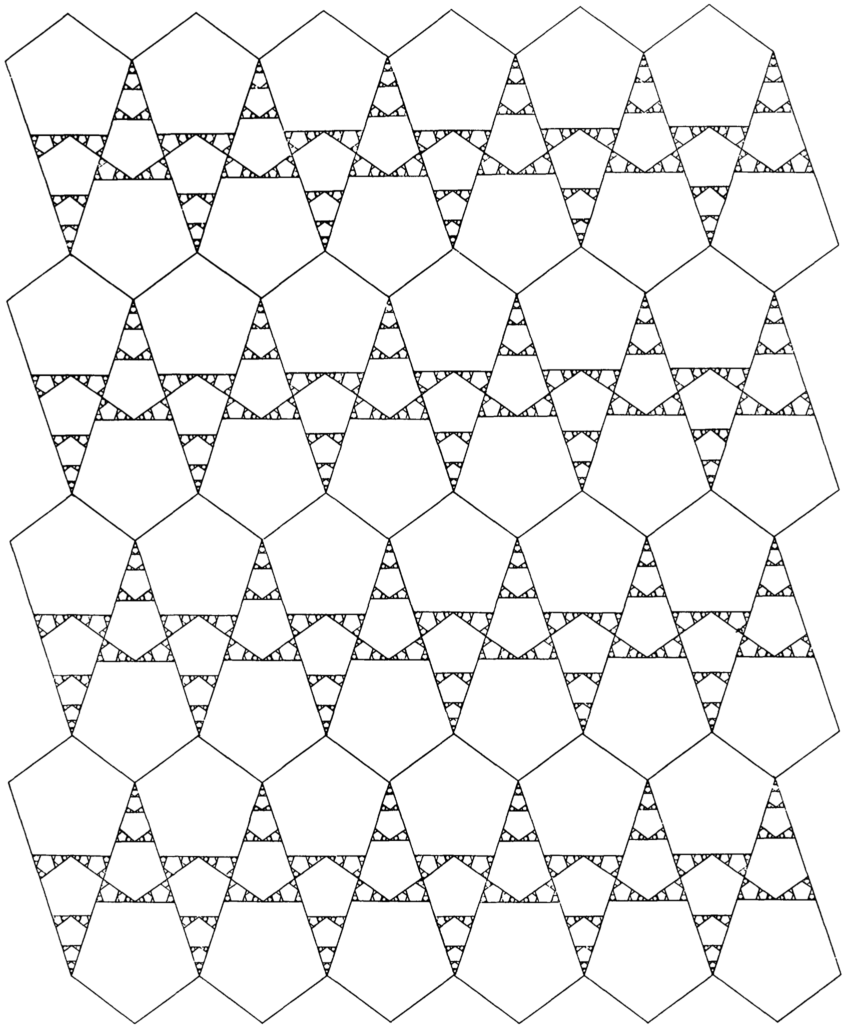


FIG. 4. A tiling which appears to consist of regular pentagons (the triangular spaces being "filled" by smaller and smaller pentagons). However, it can be shown that the union of the pentagons is a set of Cantor type, which does not cover the plane.

obtained. Similarly, the patch in Fig. 1c can be extended (by adding ten pentagons) to a patch with $\rho \approx 2.6$, while deleting from Fig. 1a the tiles which do not have 5-fold symmetry leads to a patch with $\rho \approx 3.9$.

The largest known patch made up of tiles which are regular pentagons, decagons, and pentagrams is shown in Fig. 5. Here $\rho = \sqrt{(83 + 37\sqrt{5})/8} \approx 4.6$. Patches with larger values of ρ can be constructed using the process of "decomposition"—a method which has proved useful in other tiling problems (see, for example, [**10**, Chapter 10]). This may be explained as follows. Let the edge-length of the small pentagons in Fig. 5 be denoted by s, and let each vertex of this patch be the center of a regular decagon with edge-length s/τ^3, where $\tau = \frac{1}{2}(1 + \sqrt{5})$ is the golden section ratio. Then the parts of the tiles outside these decagons can be partitioned into smaller tiles, each with 5-fold symmetry. In Fig. 6 we show this process applied to (one tenth of) the patch of Fig. 5. To obtain the whole patch we have to mirror the sector shown in the dotted lines and

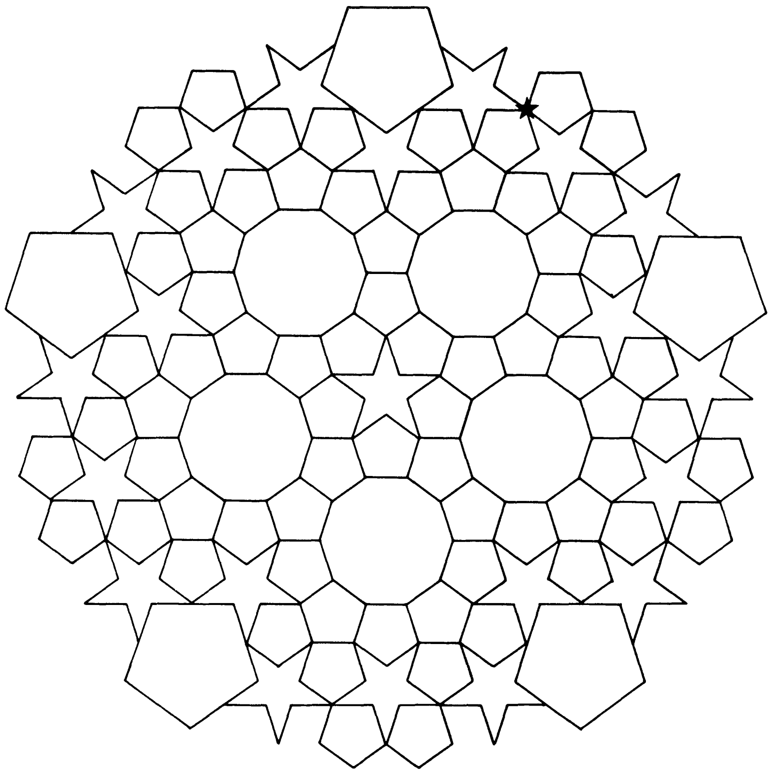


FIG. 5. A patch with the largest known $\rho \approx 4.55157\ldots$ among patches consisting of regular pentagons, decagons and pentagrams. The largest circle covered by the patch passes through the point marked by an asterisk and through its nine homologues.

their images. The original (larger) tiles are indicated by thin lines, and the tiles obtained by the decomposition process are shown by thicker lines.

Decomposition increases ρ for two reasons: the diameters of the tiles are decreased and also the patch can be extended a little using the smaller tiles. The patch in Fig. 6 has $\rho \approx 8.6$.

Repeating the process of decomposition leads to patches with larger values of ρ; in this way we can obtain patches with up to $\rho \approx 13$. Using other starting patches, we can get as high as $\rho \approx 38$. We do not know whether this is anywhere near the maximum; indeed, we cannot even prove the existence of a value $\omega < \infty$ such that $\rho \leqslant \omega$ for each patch of the kind under consideration. We conjecture that no such ω exists.

Similar problems can be raised concerning tilings and patches in which each tile has k-fold symmetry for some $k \geqslant 7$, or for any combination of such tiles and tiles with 5-fold symmetry.

References

1. E. Bindel, Harmonien im Reiche der Geometrie in Anlehnung an Keplers "Weltharmonik," Verlag Freies Geistesleben, Stuttgart, 1964.

2. J. Bourgoin, Les éléments de l'art arabe: le trait des entrelacs, Firmin-Didot et Cie., Paris, 1879. Paperback reprint of the diagrams: Arabic Geometrical Pattern and Design, Dover, New York, 1973.

3. M. J. Buerger, Elementary Crystallography, Wiley, New York, 1956.

4. M. Caspar, Welt-Harmonik, Oldenbourg, München, 1939.

5. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, 1961.

6. K. Critchlow, Islamic Patterns, An Analytical and Cosmological Approach, Schocken Books, New York, 1976.

7. S. Eberhart, New and old problems, Mathematical-Physical Correspondence, 12(1975) 4–5; *ibid.* 13(1975) 22–24.

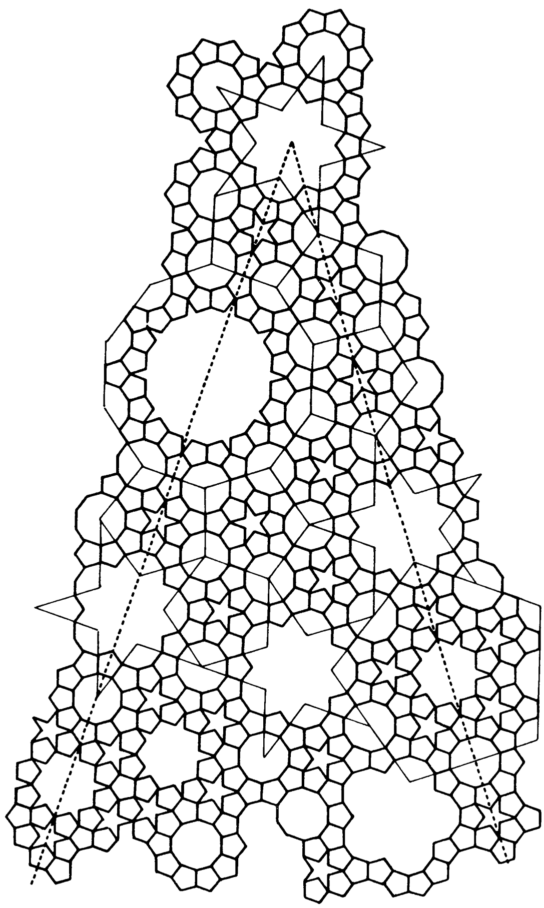


FIG. 6. A patch with $\rho \approx 8.6$ obtained by "decomposition" from the tiling in Fig. 5 and addition of some tiles around the boundary.

8. I. El-Said and A. Parman, Geometric Concepts in Islamic Art, World of Islam Festival Publishing Co., London, 1976.

9. L. Fejes Tóth, Reguläre Figuren, Akadémiai Kiadó, Budapest, 1965. English translation: Regular Figures, Pergamon, New York, 1964.

10. B. Grünbaum and G. C. Shephard, Tilings and Patterns, Freeman, San Francisco, 1982.

11. J. Kepler, Harmonice Mundi. Lincii, 1619. Reprints: Culture et Civilisation, Bruxelles, 1968; Johannes Kepler Gesammelte Werke, M. Caspar, editor, Band VI, Beck, München, 1940. Annotated German translation: M. Caspar, Welt-Harmonik, Oldenbourg, München, 1939. An annotated English translation of Book II appears in: J. V. Field, Kepler's star polyhedra, Vistas in Astronomy, 23(1979) 109–141.

12. D. Schattschneider, Tiling the plane with congruent pentagons, Math. Mag., 51(1978) 29–44.

13. ______, In praise of amateurs, The Mathematical Gardner, edited by D. A. Klarner, Prindle, Weber & Schmidt, Boston, 1981, pp. 140–166 + 5 plates in color.

14. D. Wade, Pattern in Islamic Art, Overlook Press, Woodstock, NY, 1976.

MISCELLANEA

83. Mathematics, like Dialectics, is an organ of the inner, higher mind; in practice, it is an art like eloquence. In both, nothing counts but the form; the content is irrelevant.

—J. W. v. Goethe, *Maximen und Reflexionen*, no. 605.

C E N T E R   S E C T I O N
(Vol. 89, No. 8, October 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, P. Transactions of the Moscow Mathematical Society, 1981, Issue 2. Ed: Ben Silver. AMS, 1982, iv + 284 pp, $91 (P). Translation of Tom 40 (1979), containing seven papers. LAS

General, T(13-14: 1, 2). Mathematics of Finance, Sixth Edition. Robert Cissell, Helen Cissell, David C. Flaspohler. Houghton Mifflin, 1982, xv + 569 pp, $22.95. [ISBN: 0-395-31692-8] This new edition has some new examples, some new tables to allow for current unprecedentedly high interest rates, and new problems for solution by calculator. All applications requiring calculus have been removed. (Third Edition, TR, April 1969; Fourth Edition, TR, March 1973; Fifth Edition, TR, December 1978.) FLW

General. Dictionary of Mathematics in Four Languages: English, German, French, Russian. Günther Eisenreich, Ralf Sube. Elsevier North-Holland, 1982, $174.25 set [ISBN: 0-444-99706-7]. Volume 1, 923 pp, Volume 2, 529 pp. A monumental four-language dictionary in which terms are given a subject classification. Contains about 35,000 terms. A remarkable work which should help standardize translations. SG

Elementary, T?(13: 1, 2). Algebra for College Students. Gary L. Peterson. Wadsworth Pub, 1982, x + 443 pp, $19.95. [ISBN: 0-534-00986-7] Emphasis on "problem solving," which means primarily formal algebraic manipulation, i.e., symbol pushing. For example, polynomials are introduced as sums of "products" of constants and variables. Apparently "product" is not to be thought of as a mathematical operation but as the act of juxtaposing symbols. And so it goes. Problem sets include zillions of drill problems and not much else (except in the "word problem" sections). GHM

Elementary, T(13: 1). Elementary Mathematical Methods, Second Edition. Donald D. Paige, Diane Thiessen, Margaret Wild. Wiley, 1982, viii + 413 pp, $20.95. [ISBN: 0-471-09063-8] Approaches mathematics methods with a strong laboratory orientation. Most attractive features are its focus on problem-solving skills, the use of a variety of models to develop concepts and algorithms, attention to student error patterns, and an excellent collection of problem-activities. The appendix on mainstreaming is brief but of value due to its list of references. JJ

Mathematics Appreciation, T(13-16: 1, 2). Faces of Mathematics: An Introductory Course for College Students, Second Edition. A. Wayne Roberts, Dale E. Varberg. Harper & Row, 1982, xiii + 492 pp, $21.50. [ISBN: 0-06-045471-7] An imaginative text for liberal arts students, featuring problems--creative, entertaining, challenging--and their solutions. This new edition features problems for calculator solution and a new Chapter 2 that contains review material originally located in an appendix. (First Edition, TR, August-September 1978.) LAS

Precalculus, T(13: 1, 2). Principles of Mathematics, Fourth Edition. Paul K. Rees, Charles Sparks Rees. Prentice-Hall, 1982, xvi + 508 pp, $20.95. [ISBN: 0-13-709691-7] In contrast to the previous editions, this text includes a short discussion of a "limit" prior to its use of the limit notation in the development of the concept of a derivative. Other changes involve minor reorganizations of content and the addition of extended topics in probabiity, statistics, and the integral calculus. (TR, Second Edition, February 1972; Third Edition, October 1977. JJ

Precalculus, T(13: 1), L. Fundamentals of College Algebra. Charles D. Miller, Margaret L. Lial. Scott, Foresman, 1982, 470 pp, $18.95. [ISBN: 0-673-15613-3] A competent, careful, and traditional treatment of the standard college algebra material, together with a few additions such as matrices and linear programming. JS

Precalculus, T(13: 1, 2). Algebra and Trigonometry: Precalculus Mathematics. Bodh R. Gulati, Helen Bass. Allyn & Bacon, 1982, xviii + 748 pp, $22.95. [ISBN: 0-205-07686-6] For students with two years of high school algebra. Emphasizes practical skills and computational techniques. For one or two semesters depending on student need for review, which is provided. Numerous examples and graded

exercises. Nicely laid out. Easy to read. Study guide and teacher's manual are available. JK

Precalculus, T(13), S. College Algebra. Justin J. Price, Harley Flanders. Saunders Coll Pub, 1982, xiii + 369 pp, $20.95 [ISBN: 0-03-060128-0]; College Algebra & Trigonometry. 1982, xiii + 512 pp, $21.95. [ISBN: 0-03-060132-0] Texts covering the standard topics in algebra and trigonometry. Each contains appendices on computation, including uses of calculators. JG

Precalculus, T(13: 1). College Algebra, Fifth Edition. Gordon Fuller, Walter L. Wilson, Henry C. Miller. Brooks/Cole Pub, 1982, xix + 414 pp, $20.95. [ISBN: 0-5340-1138-1] After relatively minor, largely technical revisions this Fifth Edition "retains the conservative approach of the previous editions" and still is a sound, reliable basic algebra text. (TR, Third Edition, March 1975.) JS

Education, P. Mathematics Project Handbook. Adrien L. Hess. NCTM, 1982, i + 47 pp, $3.25 (P). [ISBN: 0-87353-191-4] Revised edition of a text aimed at helping teachers guide secondary students in the selection or development of mathematical projects. Increased attention to calculators, women in mathematics, and amateur mathematicians. (TR, First Edition, January 1978.) JJ

Education, P. The High School Mathematics Library. William L. Schaaf. NCTM, 1982, ix + 84 pp, $7.80 (P). [ISBN: 0-87353-190-6] Seventh edition of NCTM's guide in the selection of books for a mathematics library. Increased emphasis on the areas of computers, calculators, and professional texts for the mathematics educator. JJ

Education, P. The Real World and Mathematics. Hugh Burkhardt. Blackie Pub, 1981, vii + 189 pp, $13.95 (P). [ISBN: 0-216-91084-6] Attempts to provide a "tool-kit" of mathematical skills to aid in recognizing, modelling, and attacking problems from the real world of secondary students. Complementary course materials and methods for the teaching of these skills are also explored and suggested. The final chapter illustrates the use of the techniques with numerous interesting applications. JJ

Education, P. Mathematics Contests: A Handbook for Mathematics Educators. David R. Johnson, James R. Margenau. NCTM, 1982, ii + 94 pp, $5 (P). [ISBN: 0-87353-187-6] Discusses the goals and organization of mathematics competitions, not only for talented students but also for students of all ability levels. Samples of award certificates, invitations, rules, and contest questions are included with a list of existing mathematics contests, field days, and leagues throughout the United States. JJ

Discrete Mathematics, S(17-18), P*. Studies on Graphs and Discrete Programming. Ed: P. Hansen. Math. Stud., V. 59. Elsevier North-Holland, 1981, viii + 395 pp, $83 (P). [ISBN: 0-444-86216-1] A varied collection of research papers relating graph theory and discrete programming. Concerns are with both theory and applications. Topics include many aspects of integer programming, assignment problems, scheduling, solving non-separable non-convex programs and applications of complexity theory. SS

Algebra, T(16-17: 2), S, L. Finite Groups and Finite Geometries. T. Tsuzuku. Transl: A. Sevenster, T. Okuyama. Tracts in Math., V. 78. Cambridge U Pr, 1982, xi + 328 pp, $42.50. [ISBN: 0-521-22242-7] A self-contained textbook relating finite groups and finite geometries. Author's goal is characterization of projective transformation groups as permutation groups. Reaches this goal with good motivation. Unfortunately, there are no exercises. SS

Algebra, S(18), P. Algebraic Structures and Applications. Ed: Phillip Schultz, Cheryl E. Praeger, Robert P. Sullivan. Lect. Notes in Pure & Appl. Math., V. 74. Dekker, 1982, ix + 168 pp, $29.50 (P). [ISBN: 0-8247-1570-5] Contains twelve of the fifteen papers presented at the First Western Australian Conference on Algebra held at the University of Western Australia in June, 1981. CEC

Algebra, T(13: 1). College Algebra. John R. Durbin. Wiley, 1982, ix + 506 pp, $21.95. [ISBN: 0-471-03368-5] Thorough review of basic mathematical concepts and skills followed by a presentation of the standard topics covered in a college algebra class. Sparse inclusion of calculator exercises. JJ

Algebra, T(18: 1), S, P. Triangular Products of Group Representations and Their Applications. Samuel M. Vovsi. Progress in Math., V. 17. Birkhauser Boston, 1981, ix + 127 pp, $10. [ISBN: 3-7643-3062-7] Beginning with Plotkin there has been a growing body of work concerned with applications of a triangular product construction to the representation of groups. The aim here is to pull much of this together in systematic fashion and to stimulate further research. Mostly deals with groups, but there is a brief appendix on associative and Lie algebras. Bibliography, no index. JS

Finite Mathematics, T(13), S. Mathematics: A Modeling Approach. Marvin L. Bittinger, J. Conrad Crown. Addison-Wesley, 1982, x + 709 pp, $22.95. [ISBN: 0-201-03116-7] A text in finite mathematics and calculus designed for students in business, economics, biology and the social sciences who have already learned basic algebra. Topics include: linear equations, linear programming, probability and statistics, decision theory, mathematics of finance, differential and integral calculus. JG

Calculus, T(13-14: 3). Calculus and Analytic Geometry, Third Edition. Sherman K. Stein. McGraw-Hill, 1982, xvi + 1211 pp, $30.95. [ISBN: 0-07-061153-X] Substantial organizational changes from Second Edition, including earlier treatment of limits (with optional ε-δ), antiderivatives and definite integrals. Many expanded sections, some deletions. A text well worth considering. (First

Edition, TR, November 1973; ER, February 1976; Second Edition, TR, May 1977.) GHM

Calculus, T(13), S*. Calculus by Discovery. Douglas Downing. Barron's Educ Ser, 1982, xii + 273 pp, $12.95. [ISBN: 0-8120-5451-2] Entertaining fantasy/adventure in which the ideas and methods of calculus are developed through the joint efforts of the characters in the story. Demonstrations are informal, but generally complete and convincing. Text has advantage of answering students' questions as they arise. JRG

Calculus, T(13: 2). Calculus and Analytic Geometry. Abe Mizrahi, Michael Sullivan. Wadsworth Pub, 1982, xviii + 1112 pp, $34.95. [ISBN: 0-534-00978-6]; Study Guide. Richard St.André. 249 pp, (P). The usual topics and physical science applications are attractively presented in a conversational style which makes this text very readable but somewhat lengthy; theorems are clearly stated with many proofs relegated to an appendix. The Study Guide contains objectives, synopses and diagnostic quizzes for Chapters 2-13 of the nineteen-chapter text. JNC

Calculus, T(13-14: 1-3). Applied Calculus for Business and Economics, Life Sciences, and Social Sciences. Raymond A. Barnett, Michael R. Ziegler. Dellen Pub, 1982, xiv + 674 pp, $23.95. [ISBN: 0-89517-036-1] For students with one and one-half to two years of high school algebra. Review of same in Chapter 1 and Appendix. Trigonometry, required for Chapter 7, is not reviewed. Emphasizes computational skills and problem solving. Slights derivations and proofs. Includes multivariable calculus, differential equations, numerical techniques and probability. Unfortunately, most applications are of the type "The formula for...is... . Find... ." Graded exercise sets. Solutions manual and instructor's manual available. JK

Calculus, S(14-16), L. Mathematical Methods for Mathematicians, Physical Scientists, and Engineers. J. Dunning-Davies. Math. and Its Appl. Halsted Pr, 1982, 416 pp, $57.95. [ISBN: 0-470-27322-4] Summary of useful techniques from calculus of one and several real variables, complex analysis, linear algebra, vector and tensor analysis, ordinary and partial differential equations, Fourier series and special functions. Statistics and numerical methods are omitted. Not suitable for text in U.S.A. Except for cost, a handy reference. Some exercises, most with answers. JK

Calculus, T(13-14: 1-3), L. Calculus with Analytic Geometry, Second Edition. Roland E. Larson, Robert P. Hostetler. DC Heath, 1982, xxi + 1081 pp, $31.95. [ISBN: 0-669-04530-6] A more thorough revision than many second editions, this one includes the addition of new chapters on line integrals and differential equations as well as some "substantial changes in exposition" in a number of individual sections. A good book that seems improved. (First Edition, TR, June-July 1979.) JS

Real Analysis, T*(14-15: 1), S. Introduction to Real Analysis. Robert G. Bartle, Donald R. Sherbert. Wiley, 1982, xii + 370 pp, $28.95. [ISBN: 0-471-05944-7] Topics are those of elementary calculus, with some additions. Pace and rigor appropriate to an intermediate course. Covers, e.g., limits, continuity, the Riemann integral, convergence of functions (pointwise and uniform), some topological ideas, all in one real-variable, all with examples. Well-motivated by freshman calculus, nicely motivates advanced calculus. Many exercises, hints. PZ

Real Analysis, P. Theory and Applications of Differentiable Functions of Several Variables. VII. Ed: S.M. Nikol'skii, E.A. Volkov. Proc. of Steklov Inst. of Math., 1981, Issue 4. AMS, 1982, vii + 336 pp, $88 (P). [ISBN: 0-8218-3017-1] Translation of Volume 150, 1979, containing 16 papers in a single area. The seventh in a series of Steklov papers on differentiable functions. LAS

Complex Analysis, T(17), S, P. Introduction to Complex Analysis, Second Edition. Rolf Nevanlinna, V. Paatero. Transl: T. Kövari, G.S. Goodman. Chelsea Pub, 1982, ix + 350 pp, $14.95. [ISBN: 0-8284-0310-4] Graduate level introductory text emphasizes analytic functions as mappings--properties of elementary functions and associated Riemann surfaces are thoroughly developed. Contains: homotopy of contours, harmonic measure, boundary correspondence under conformal mapping, infinite products, special functions. Omits: approximation by rational functions (Runge's theorem). PZ

Complex Analysis, T(15-16: 1), S. Basic Complex Variables for Mathematics and Engineering. John H. Mathews. Allyn & Bacon, 1982, xii + 319 pp, $24.95. [ISBN: 0-205-07170-8] Mathematical development is relatively elementary--e.g., avoids uniform convergence. Stresses, instead, interesting applications to physics and engineering of conformal mapping, harmonic functions, residue integrals. Excellent, helpful graphics. Extraordinary number and variety of exercises, many applied. PZ

Differential Equations, T(14), S, L. Ordinary Differential Equations with Applications. Larry C. Andrews. Scott, Foresman, 1982, x + 338 pp, $22.95. [ISBN: 0-673-15800-4] A text for a first course in differential equations. Topics include: linear equations, Laplace transforms, Green's functions, systems of equations, numerical methods, power series methods. JG

Differential Equations, P. Spectral Theory of Differential Operators. Ed: Ian W. Knowles, Roger T. Lewis. Math. Stud., V. 55. Elsevier North-Holland, 1981, xv + 384 pp, $46.75 (P). [ISBN: 0-444-86277-3] Proceedings of a conference held in March, 1981, at the University of Alabama. The papers present recent work in the theory of ordinary and partial differential equations which may be considered under the general heading of spectral theory. AO

Differential Equations, P. Solitons and the Inverse Scattering Transform. Mark J. Ablowitz, Harvey Segur. Stud. in Appl. Math. SIAM, 1981, x + 425 pp, $54.50. [ISBN: 0-89871-174-6] The inverse scattering transform is a method which can be used to obtain exact solutions to certain nonlinear

evolution equations. It can be viewed as a generalization of the Fourier transform. This monograph is an exposition of the technique and some of its applications. AO

Differential Equations, P*. Ill-Posed Problems for Integrodifferential Equations in Mechanics and Electromagnetic Theory. Frederick Bloom. Stud. in Appl. Math. SIAM, 1981, ix + 222 pp, $34.50. [ISBN: 0-89871-171-1] Treatment mainly based on variants of logarithmic convexity and concavity arguments. Final chapter surveys recent research results and directions for ill-posed initial-history boundary value problems for integrodifferential equations with an emphasis on non-linear first-order equations. Prerequisites include some Hilbert space theory, basic formulae for linear partial differential equation theory, and elementary properties of convex real-valued functions of one variable. Well-referenced. Good index. JK

Differential Equations, P. Integrable Systems: Selected Papers. S.P. Novikov. London Math. Soc. Lect. Note Ser., V. 60. Cambridge U Pr, 1981, 266 pp, $27.50 (P). [ISBN: 0-521-28527-5] This volume is a collection of seven papers originally published in Russian Mathematical Surveys and an introductory essay all on the subject of 'integrable' non-linear partial differential equations. AO

Differential Equations, T*(15-16: 1), S*, L.** Ordinary Differential Equations: A Qualitative Approach with Applications. D.K. Arrowsmith, C.M. Place. Math. Ser. Chapman & Hall, 1982, ix + 252 pp, $15.95 (P); $29.95. [ISBN: 0-412-22610-3; 0-412-22600-6] This text is appropriate for use in an introductory course on the qualitative theory of ordinary differential equations for students who have studied multivariable calculus and elementary linear algebra. Applications of the theory are integrated throughout the text and range from mechanical and electrical systems to models in ecology, economics and cancer research. AO

Numerical Analysis, P. The Solution of Initial Value Problems Using Interval Arithmetic: Formulation and Analysis of an Algorithm. P. Eijgenraam. Math. Centre Tracts, V. 144. Math Centrum, 1981, iii + 185 pp, Dfl. 24,15 (P). [ISBN: 90-6196-230-7] Develops a new method for obtaining interval solutions at discrete time steps for a system of ordinary differential equations for which an interval for the initial value is given. Global analysis. Program. Numerical results. RWN

Numerical Analysis, T(15-16: 1), L. Numerical Methods: A Software Approach. R.L. Johnston. Wiley, 1982, ix + 276 pp, $24.95. [ISBN: 0-471-09397-1] An introduction to numerical algorithms used in high-quality mathematical software in the context of a first course in numerical analysis: linear algebra, interpolation, nonlinear equations, quadrature, and differential equations. A careful compromise between traditional, full-fledged numerical analysis and common (albeit dangerous) black-box use, designed to encourage intelligent use of mathematical software packages. LAS

Functional Analysis, S(18), P. Semigroup and Factorization Methods in Transport Theory. C.V.M. van der Mee. Math. Centre Tracts No. 146. Math Centrum, 1981, ii + 167 pp, Dfl. 21 (P). [ISBN: 90-6196-233-1] A treatment of the transport equation for two problems--the finite slab and the half-space problem. Several techniques are developed including use of semi-groups of operators, Weiner-Hopf factorization, and recent work of Gohberg and Heinig. Bibliography, index. JS

Functional Analysis, P. Interpolation of Linear Operators. S.G. Krein, Ju. I. Petunin, E.M. Semenov. Transl. of Math. Mono., V. 54. AMS, 1982, xii + 375 pp, $71.20. [ISBN: 0-8218-4505-7] This book presents some of the main ideas in the theory of interpolation of linear operators: the real and complex methods of constructing interpolation spaces, the method of scales of Banach spaces, and interpolation in spaces of measurable functions. These ideas allow one to view many results of classical analysis from a new perspective. AO

Analysis, P. Contributions to Analysis and Geometry. Ed: D.N. Clark, G. Pecelli, R. Sacksteder. Johns Hopkins U Pr, 1981, ix + 357 pp, $36.50. [ISBN: 0-8018-2779-5] Papers from an April 1980 conference at Johns Hopkins, published as a supplement to the American Journal of Mathematics. LAS

Algebraic Geometry, P. Birational Geometry for Open Varieties. Shigeru Iitaka. Pr U Montreal, 1981, 94 pp, $11 (P). [ISBN: 2-7606-0561-2] An introduction to birational algebraic geometry similar to the author's recent Springer Graduate Text. Among the topics: D-dimension, Kodaira dimension, logarithmic forms, and automorphism groups. SG

Differential Geometry, P. The Exponential Map at an Isolated Singular Point. David A. Stone. Memoirs No. 256. AMS, 1982, v + 184 pp, $10.40 (P). A study of how the calculus of variations can be used to study geodesics in manifolds with singularities. Applications to complex analytic hypersurfaces. JG

Differential Geometry, S(18), P. Analysis on Topological Groups--General Lie Theory. Helmut Boseck. Günter Czichowski, Klaus-Peter Rudolph. BG Teubner, 1981, 136 pp, 14M (P). With an announced goal of a generalized differential calculus for infinite-dimensional Lie groups, foundations are developed for a category of topological groups (pre-Lie groups). Assumes the set of one-parameter subgroups is a topological Lie algebra and makes extensive use of the exponential map and the Campbell-Hausdorff formula. References, index. JS

Differential Geometry, P. Synthetic Differential Geometry. Anders Kock. London Math. Soc. Lect. Note Ser., V. 51. Cambridge U Pr, 1981, 311 pp, $27.50 (P). [ISBN: 0-521-24138-3] Intended to lay the foundation of "synthetic reasoning" in differential geometry; "space forms" are the objects of a category whose morphisms are the constructions which are performed on the forms. The basic category

is a Grothendieck topos. SG

Algebraic Topology, T(16-18: 1, 2), S. Algebraic Topology, A First Course. Marvin J. Greenberg, John R. Harper. Math. Lect. Note Ser. Benjamin/Cummings Pub, 1981, xi + 311 pp, $19.50 (P); $31.50. [ISBN: 0-8053-3557-9; 0-8053-3558-7] Revision of Lectures on Algebraic Topology by the first author. Additions have been made to the theory, examples, exercises, calculations, and geometry. Text includes four parts: elementary homotopy theory; singular homology theory; orientation and duality on manifolds; products and the Lefschetz fixed point theorem. Bibliography. Index. RJA

Topology, S(18), P. Finite Group Actions on Simply-Connected Manifolds and CW Complexes. Amir H. Assadi. Memoirs No. 257. AMS, 1982, x + 116 pp, $7.20 (P). This monograph looks at the existence and construction of embeddings of a given G-CW complex (G-manifold) in another G-CW complex having a prescribed homotopy type and a prescribed family of isotropy subgroups. Based on the author's doctoral dissertation. CEC

Statistics, T(13-14: 1, 2). Basic Statistics for Business and Economics, Third Edition. Paul G. Hoel, Raymond J. Jessen. Wiley, 1982, x + 629 pp, $23.95. [ISBN: 0-471-09829-9] This new edition has some new examples, a new chapter on non-parametric methods, and additional material on multiple regression, time series, and decision theory. (TR, First Edition, March 1973; Second Edition, November 1977.) FLW

Statistics, T(13-16: 1, 2), S, L. Statistical Analysis of Geological Data. George S. Koch, Jr., Richard F. Link. Dover, 1980, ix + 438 pp, $12.50 (P). [ISBN: 0-486-64040-X] Presupposes only elementary algebra and geometry. Emphasizes and illustrates the methods most useful in geology: univariate methods, analyses of variance, variable transformation, and multivariate methods. No exercises. FLW

Statistics, T(13-14: 1, 2). Statistics by Example. Terry Sincich. Dellen Pub, 1982, xiii + 578 pp, $24.95. [ISBN: 0-89517-037-X] The usual topics, plus multiple regression, with the discussions built around real data. Presupposes only high school mathematics. Some examples of the use of computer packages included. FLW

Statistics, S*(14-18), P, L*. Encyclopedia of Statistical Sciences, Volume 1: A to Circular Probable Error. Ed: Samuel Kotz, Norman L. Johnson. Wiley, 1982, x + 480 pp, $75. [ISBN: 0-471-05546-8] First volume of a planned eight-volume set designed "to provide information about an extensive selection of topics concerned with statistical theory and the applications of statistical methods in various more or less scientific fields of activity." Includes references to the literature and cross-references to related entries. An invaluable source for the non-specialist. RSK

Statistics, P*. Contributions to Statistics. William G. Cochran. Wiley, 1982, xviii + 1833 pp, $85. [ISBN: 0-417-09786-1] In the Wiley Series in Probability and Mathematical Statistics. Collection of all 116 of Cochran's published articles, written between 1934 and 1980, arranged in chronological order. Of particular interest to anyone interested in the history of statistics. No index. RSK

Statistics, T*(13-14: 1). Business Statistics by Example. Terry Sincich. Dellen Pub, 1982, xiii + 658 pp, $23.95. [ISBN: 0-89517-035-3] An intriguing introduction to statistics as motivated by 49 actual case studies. Four extensive data sets are also included and are used to motivate concepts such as sampling distributions or population parameters. The problem sets are interesting and complement the text's approach. JJ

Statistics, T(13-14: 1, 2). Statistics: Discovering Its Power. Ronald J. Wonnacott, Thomas H. Wonnacott. Wiley, 1982, xviii + 378 pp, $20.95. [ISBN: 0-471-01412-5] In Wiley's Probability and Mathematical Statistics Series. Uses examples to intuitively motivate statistical concepts at "an easy and applied level." Focuses on regression analysis and estimation techniques using confidence intervals and p-values rather than on classical hypothesis testing. Problem sets are of graded difficulty and are complemented by the use of appropriate computer packages. JJ

Statistics, T(13-14: 1). Elements of Statistical Reasoning. Edward W. Minium, Robert B. Clarke. Wiley, 1982, xv + 470 pp, $18.95. [ISBN: 0-471-08041-1] Presupposes only high school mathematics. Basic mathematics is reviewed in the appendix. The usual topics including analysis of variance. FLW

Statistics, T(16-18: 1, 2), S, P, L. Acceptance Sampling in Quality Control. Edward Schilling. Statistics, V. 42. Dekker, 1982, xx + 775 pp, $65. [ISBN: 0-8247-1347-8] A compendium of acceptance sampling methods. Many useful tables and graphs. Presupposes a course in quality control. A student price of $37.50 is available if 5 or more copies are ordered. FLW

Computer Science, S(14-16), L. Draft Proposal for the B-Programming Language: Semi-Formal Definition. Lambert Meertens. Math Centrum, 1981, 88 pp, Dfl. 11,55 (P). [ISBN: 90-6196-238-2] Contains a description of B which is intended to be a simple but powerful language to be used on personal computers for applications such as record keeping, scientific computations and education. Some innovations in the language should be of general interest. RWN

Applications (Biology), S(15-17), L*. Mathematical Methods of Population Biology. Frank C. Hoppensteadt. Stud. in Math. Biology, V. 4. Cambridge U Pr, 1982, viii + 149 pp, $29.95; $12.95 (P). [ISBN: 0-521-23846-3; 0-521-28256-X] This short book is an introduction to some of the mathematical methods useful in studying models of population dynamics formulated in terms of difference equations. The mathematical methods include elementary differential equations, Markov chains, perturbation techniques, and partial differential equations. May be particularly useful as a source of illustrative examples. AO

Applications (Biology), S, L*. Applicable Mathematics of Non-Physical Phenomena. F. Oliveira-Pinto, B.W. Conolly. Math. & Its Appl. Halsted Pr, 1982, 269 pp, $59.95. [ISBN: 0-470-27297-X] An anthology of hard-to-find seminal twentieth century papers in biological and medical applications: Volterra, Kolmogorov, Feller, Hardy, and others. Each paper is introduced with sufficient material to establish an appropriate mathematical and notational context. Could serve as an interesting source for a modelling seminar. LAS

Applications (Economics), S(16-18), P, L.** Handbook of Mathematical Economics, V. II. Ed: Kenneth J. Arrow, Michael D. Intriligator. Elsevier North-Holland, 1982, xix + 688 pp, $150 set of 3 volumes. [ISBN: 0-444-86126-2] Second of three volumes surveying the state-of-the-art of mathematical economics. A masterful exposition of current knowledge by leading experts, this volume comprises Chapters 9-21, covering microeconomic theory and competitive equilibrium. Each chapter can be read independently and contains extensive references to current literature. (Volume I, TR, March 1982.) LAS

Applications (Engineering), T*(15-16: 1, 2), S, L. Advanced Engineering Mathematics. Ladis D. Kovach. Addison-Wesley, 1982, xi + 706 pp, $23.95. [ISBN: 0-201-10340-0] Standard techniques used in engineering analysis. For physics and mathematics students as well. Flexible. Useable for brief, separate courses in ordinary differential equations, linear algebra, vector analysis, complex variables and boundary-value problems. Some numerical methods. Stress on applications. Over 2000 exercises in three levels--routine, practice, theoretical-challenging. Many references, some off the beaten track. JK

Applications (Physics), S(17-18), P*. Angular Momentum in Quantum Physics: Theory and Application. L.C. Biedenharn, J.D. Louck. Encyclopedia of Math. & Its Appl., V. 8. Addison-Wesley, 1981, xxxii + 716 pp, $54.50. [ISBN: 0-201-13507-8] A comprehensive treatment of angular momentum theory and some of its applications to physical problems. In addition to the standard theory, it includes much that is new. AO

Applications (Physics), P. Quantum Gravity 2: A Second Oxford Symposium. Ed: C.J. Isham, R. Penrose, D.W. Sciama. Clarendon Pr, 1981, xiv + 669 pp, $39.95. [ISBN: 0-19-851952-4] A review in April 1980 of developments since the first (1974) Oxford Symposium on quantum gravity on one of the major unattained goals of present day theoretical physics--the union of quantum theory with Einstein's general theory of relativity. Extensive survey articles make this an ideal introduction for researchers and students of quantum physics. LAS

Applications (Physics), S(14-16). Newtonian Attraction. A.S. Ramsey. Cambridge U Pr, 1981, ix + 184 pp, $14.95 (P). [ISBN: 0-521-09193-4] A re-issue of the 1940 original, this text provides an introduction to gravitational potential theory. AO

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; JJ: Jerry Johnson, St. Olaf; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Southwestern Section

The Southwestern Section held its annual meeting in conjunction with La Sociedad Matematica de Sonora at the University of Arizona in Tucson, Arizona, Friday and Saturday, April 2-3, 1982. Seventy-one members registered for the meeting.

Invited Addresses:

"True Facts About Imaginary Objects," by Rueben Hersh, University of New Mexico.
"Drama in Mathematics: A Demonstration with Partitions," by Henry Alder, University of California, Davis.
"A Scholar's Trip to China," by John Brillhart, University of Arizona.

Short Presentations:

"Spinors in a Geometrical Mode," by Alvin Swimmer, Arizona State University.
"Ptolemy's Cosmology," by Steve Hammel, University of Arizona.
* "Euclidean Construction of the Foci and Directrices of a Given Ellipse Via the Dandelin Spheres," by Charles G. Moore, Northern Arizona University.
"Computations in Finite Fields Using Lech's Logarithm," by Peter Tennenbaum, University of Arizona.
"An Improvement on Liouville's Theorem for Quadratic Irrationals," by Everett Walter, Northern Arizona University.
"A Curious Characterization of the Exponential Function," by Louis A. Talman, College of Ganado.
* "On the Order of a Banach Space," by Ray Griesan, University of Arizona.
"On a Theorem of Van der Waerden," by Daniel J. Madden, University of Arizona.
"Mathematics and the Sheriff's Department: A Simple Bayesian Approach to Search and Rescue in the Sonoran Desert," by John Bownds, University of Arizona.
"Tetrahedral Kaleidoscopes: The Mathematical Gem and Mineral Show Explained," by Tom Weaver, University of Arizona.
* "Humanizing the Mathematics Classroom," by Eugenia E. Fitzgerald, Phoenix College.
* "Comparison of the Performance of the University Freshman Who Graduated High Schools in Mexico and in the United States and Attended the University of Texas at El Paso," by J.R. Provencio, University of Texas at El Paso.
"Lattice-ideals in Lattice-ordered Groups," by Justin G. MacCarthy, Deming, New Mexico.
"The Mobius Inversion Formula for Arithmetic Functions," by Michael I. Ratliffe, Northern Arizona University.
* "The Lattice of Subfields of a Radical Extension," by William Yslas Velez, University of Arizona.
* "An Analogue of Mutual Singularity for Measures on a Field," by John W. Hagood, University of Arizona.
"The Calculation of Lower Bounds to Eigenvalues of the Atomic Schrodinger Equation," by David M. Russell, University of Arizona.
* "Punto fijos y ecuaciones diferenciales (Fixed Points and Differential Equations)," by Rueben Flores, Universidad de Sonora.
"Reaction-diffusion Equations, One Place that O.D.E.'s and P.D.E.'s Meet," by Chris Jones, University of Arizona.
* "Variational Principles and Minimum Energy Equilbrium Solutions in Physical Field Theories," by Hanno Rund, University of Arizona.
"Running a Calculator Workshop," by Lee Erlebach, Michigan Technological University.
"Progress Report on Calculus Using a Computer Based Instruction," by David Lomen, University of Arizona.

Special Presentation:

"Projecto de Maestria en mathematical educativa (Master's Program in Mathematics Education)," by Marco Antonio Valencia, Universidad de Sonora.

Panel Discussions:

"Better Mathematical Preparation for Entering College Students: Some Exciting Recent Developments," by Henry Alder, University of California, Davis; David Gay, University of Arizona; Diana Bishop, Tucson Unified School Districts; Lou Smith, Burr-Brown; Alejandro Lopez Yanez, Universidad Nacional Autonoma de Mexico.
"The Use of the Computer in Teaching of Differential Equations," by Tom Kyner, University of New Mexico; David Lomen, University of Arizona; Jim Cushing, University of Arizona.

New Jersey Section

The spring meeting of the New Jersey Section was held on Saturday, April 17, 1982 at Georgian Court College, Lakewood, New Jersey. Fifty people registered for the meeting, which was held in conjunction with the Mathematics Association of Two-Year Colleges of New Jersey.

Invited Addresses:

"Emmy Noether, Her Life," by Gottfried Noether, University of Connecticut, and Patricia Kenschaft, Montclair State College.
"Emmy Noether, Her Work," by Mark Krusemeyer, Rutgers University.
"Worrisome Wagers: Intuitive or Counterintuitive Mathematics?" by Warren Page, New York City Technical College (CUNY).
"Soap Bubble Geometry," by Jean Taylor, Rutgers University.

Panel Discussion:

"CUPM's Mathematical Sciences Program: Implications for Two- and Four-Year Colleges." Virginia McGlone (Moderator), College of St. Elizabeth; Fred Roberts, Rutgers University; John Houle, Pace University; Monir Kashmiry, Union College; Evan Alderfer, Ocean County College.

Missouri Section

The Missouri Section held its spring meeting at the University of Missouri at Rolla on April 10, 1982.

Invited Addresses:

"Paradoxical Coverings of the Real Line," by Ivan Niven, University of Oregon.
"Educational Applications of Computer Graphics," by O.R. Plummer, University of Missouri at Rolla.
"Are You Ready for Calculus Calculators?" by Arlan DeKock, University of Missouri at Rolla.

Short Presentations:

"Taylor's Formula with Remainders," by Trent Eggleston, University of Missouri at Rolla.
"Replacement of Equality Constraints by Inequality Constraints," by Johnny Roberts, University of Missouri at Rolla.
"Exponentiation Without Associativity," by Gary Birkenmeier and Steve Plaskemeier, S.E.M.S.U.
"Mathematizing 'Frogs': Heuristics, Proof, and Generalization in the Context of a Recreational Problem," by Gary G. Cochell, Culver-Stockton College.
"Statistical Metric Spaces: Examples and Topological Classifications," by Troy Hicks, University of Missouri at Rolla.
"N.C.A.T.E. and Its Influence on Mathematics," by Don Hight, Pittsburg State University.

Illinois Section

The annual meeting of the Illinois Section convened at Southern Illinois University, Edwardsville, on Friday and Saturday, April 30-May 1, 1982 with approximately seventy members in attendance.

Invited Addresses:

"We Don't Know How To Teach Our Way Out of a Paper Bag," by Leonard Gillman, University of Texas.
"From Saturn with Parity: An Introduction to Interplanetary Coding Theory," by Robert J. McEliece, University of Illinois.
"Mathematics Teacher--Education and Certification," by John Dossey, Illinois State University.
"The p-version of the Finite Element Method," by I. Norman Katz, Washington University.
"The Importance of VE in DeVElopmental Math," by Larry Johansen, Kishwaukee College.
"Reciprocity Laws of Number Theory," by Carlos Moreno, University of Illinois.
"Ball-point, Chalk and Greasepen," by George Francis, University of Illinois.
"The Geometry of Control Systems," by David L. Elliott, Washington University.
"A Magnificent Generalization of the Pigeon-Hole Principle," by Marilyn Livingston, Southern Illinois University.
"Homogeneous Temperatures," by Deborah Tepper Haimo, University of Missouri.
"Mathematical Roots: Topics in the History of Mathematics," by John Schumaker, Rockford College.

Professor Arnold Wendt of Western Illinois University was presented the Distinguished Service Award of the Section.

Indiana Section

The spring meeting of the Indiana Section was held at Ball State University on Saturday, April 24, 1982, with approximately 60 members present. The Indiana Small College Math Competition was held in conjunction with the meeting.

Invited Address:

"Linear Algebra Made Difficult," by Paul R. Halmos, Indiana University.

Short Presentations:

"Loci with a Carpenter's Square," by Rodney T. Hood, Franklin College.
"An Extension of Dirichlet's ($m+n\alpha$) Theorem," by Richard A. Bieberich, Ball State University.
* "Shape Instabilities of Growing Bodies," by John Chadam, Indiana University.
"How Does Your Calculator Find the Sine of 13?" by Richard R. Patterson, Indiana University-Purdue University at Indianapolis.
"Splines Under Tension," by Elton G. Graves, Rose Hulman Institute of Technology.

Allegheny Mountain Section

The Allegheny Mountain Section held its annual memeting at Allegheny College in Meadville, Pennsylvania on Friday and Saturday, April 30-May 1, 1982.

Invited Addresses:

"The Ecological Distance Between Species," by Ron Harrell, Allegheny College.
"Algorithms for Generating a DeBruijn Sequence," by Anthony Ralston, SUNY at Buffalo.
"The Mathematics of M.C. Escher's Art," by Doris Schattschneider, Moravian College.
"Finding Your Way in a Graph," by Ronald Graham, Bell Labs.

Short Presentations:

"Coffee and Cancer of the Pancreas: A Statistical Review," by J. Konrad Sauer, Allegheny College.
"Percolation Theory: A New Tool for Physicists," by David Stiffler, Allegheny College.
"Computing Electrostatic Fields Using Conformal Mappings and the Schwartz-Christoffel Formula," by Carolyn Kyler, Allegheny College.
"The Group of Rubik's Cube," by Margaret McAdams, Allegheny College.
"Generated Term of a Sequence," by John W. Milsom, Butler County Community College.
"Sums of Squares in the Gaussian Integers," by John B. Lane, Edinboro State College.
"Non-Commutative Rings with Small Order," by Jim Derr, West Virginia University.

Panel Discussion:

"Mathematics Required/Desired in the Computer Science Curriculum." Donald Platte, Mercyhurst College (moderator).

North Central Section

The spring 1982 meeting of the North Central Section was held at Saint John's University, Collegeville, Minnesota, on April 23-24, 1982.

Invited Addresses:

"The Influence of Mathematical Ideas on Thomas Jefferson's Declaration of Independence," by Sylvan Burgstahler, University of Minnesota, Duluth.
"Some Bridges To and From Mathematics," by Alfred B. Willcox, Executive Director of the MAA.

Short Presentations:

* "The Third Derivative," by George Bridgman, Macalester College.
* "Let's Differentiate 1/2 Times: A Child's Garden of Generalized Derivatives," by Ravi N. Kalia, Southwest State University.
"Direct Sum Decompositions of Modules in the Generalized Noetherian Setting," by Francis T. Hannick, Mankato State University.
"The Wobbley Box," by Michael Tangredi, College of Saint Benedict.
* "A Continuous Model of Epidemic-Genetic Interaction," by John T. Kemper, College of Saint Thomas.
"The University of Minnesota's Institute for Mathematics and Its Applications," by George Sell, University of Minnesota.
* "The Number of Homomorphisms from Z_m to Z_n," by Joseph A. Gallian, University of Minnesota, Duluth.
"Counting Subgroups of a Finite Group," by William Calhoun, Carleton College.
* "How True is a Correctness Function: Image-Based Definition of Correctness," by Abraham Shammas, Southwest State University.
* "Odds Versus Evens in Silverman Games," by Gerald A. Heuer, Concordia College.
* "Some Elementary Questions About Metrics," by Harold Martin, St. Cloud State University.

Seaway Section

The Seaway Section held its spring meeting at Skidmore College, Saratoga Springs, New York on April 23 and 24, 1982. Approximately 60 people attended the meeting.

Invited Addresses:

"Improved Spatial Skills in Mathematics Using Computer Graphics," by Edwin H. Rogers, Rensselaer Polytechnic Institute.
"Computational Complexity of Feasible Computations," by Juris Hartmanis, Cornell University.

Short Papers:

"Analogous Spaces: Integrals and Complex Products," by J.B. Harkin, SUNY at Brockport.
"A Student Centered Math Skills Center: Three Years of Experience at SUNY at Geneseo," by Gilbert A. Palmer, SUNY at Geneseo.

"Weighted Voting for County Boards in New York," by David Housman, Cornell University.
"Discriminant Analysis for Patient Placement," by Martin Orr, New York Health Department.
"Mathematics and Computer Science: A Curriculum, A Point of View, and the Future," by Thomas M. O'Loughlin, SUNY at Cortland.
"Large Aerospace Structures: Do I Have to Understand Hilbert Space to Live in Outer Space?" by Mark J. Balas, Rensselaer Polytechnic Institute.
"Why Vectors Are Not Arrows," by Douglas L. Cashing, St. Bonaventure University.

At the business meeting, David W. Ash of the University of Waterloo received the Section Putnam Prize.

Rocky Mountain Section

The sixty-fifth annual meeting of the Rocky Mountain Section was held on April 30-May 1, 1982 on the campus of Western State College in Gunnison, Colorado with 114 members of the Section in attendance.

Invited Address:

"Paradoxes about Rational and Irrational Numbers," by Ivan Niven, University of Oregon.

Panel Discussions:

"Mathematics Support Courses for the Disciplines of Engineering, Natural Science and Computer Science." John Gill, University of Southern Colorado (moderator). Panelists: Bob Brown, Colorado School of Mines; Ralph Niemann, Colorado State University; and David Ballew, South Dakota School of Mines and Technology.
"Mathematics Service Courses for the Disciplines of Business, Education, and the Social Sciences." George Donovan, Metro State College (moderator). Panelists: Vern Nelson, Metro State College; Gale Nash, Western State College; and John Hodges, University of Colorado.

Short Presentations:

"The Teaching of Computing at a Distance," by P.G. Thomas, The Open University and Denver University.
"A Generalized Problem of Least Squares," by Hung-Chiang Li, University of Southern Colorado.
"Some Space Shuttle Mathematics," by Raymond E. Williams, Ft. Lewis College.
"Installing a Stove Pipe," by Aubrey P. Owen, Community College of Denver North.
"Two Computer Programming Problems," by David Ballew, South Dakota School of Mines and Technology.
"Mathematics at Work in Society," by John Jobe, Oklahoma State University.
"How to Obtain that Grant," by John Jobe, Oklahoma State University.
"An Interdisciplinary Applied Mathematics Graduate Program," by William B. Jones, University of Colorado.
"An Undergraduate Mathematics Problem Course at Colorado State University," by Arne Magnus, Colorado State University.
"Fixed Point Theorems," by V.M. Schgal, University of Wyoming.
"Sets of Integers with Integer Mean and Standard Deviation--Good Test Questions," by Edward M. Corwin, South Dakota School of Mines and Technology.
"On Lacunary Series," by S.A. Husam, University of Wyoming.
"Properties of Solutions of the Difference Equation $Y(t+1) = F(t,Y(t))$," by Gary Grefsrud, Ft. Lewis College.
"Some Topics Visited in the Analytic Theory of Continued Fractions," by W.J. Thron and Haakson Waadeland, University of Colorado.
"Shader's Circles," by Leslie E. Shader, University of Wyoming.
"Simple Iterative Constructions that Densely Fill Space," by Gary Brennen, University of Colorado.
"Beware the Bessel Function Expansion," by Albert Grimm, South Dakota School of Mines and Technology.

Student Papers:

"A Problem in Number Theory," by Boris Lerner, University of Colorado.
"Algebra and Energy," by Debbie Wentzel, South Dakota School of Mines and Technology.
"Hardy-Wienberg Equilibrium," by Susan German, Air Academy.
"Number of Roulette Wheels with n Colors--A Group Theory Approach," by Mark K. Raus, Air Academy.
"Fibonacci Sequence," by T.J. Sakulich, Air Academy.
"Programming the Khachiyan Algorithm," by Brian Bunsness, South Dakota School of Mines and Technology.
"A Simple Computer Method for Calculation of pi to n Decimal Digits in t Time," by Gregory Bollendonk, Ft. Lewis College.
"On a Problem of Collatz," by Gary P. Shea, University of Colorado.
"Mathematics at the Air Academy," by John Garstka, Air Academy.

"Simultaneous Plays of Infinite Games," by Steve Grantham, University of Colorado.
"Effective Rate of Return," by Jonathan Wade, Colorado School of Mines.
"Newton's Method for Complex Variables," by Dean Mogck, South Dakota School of Mines and Technology.
"Real Valued Bases," by David White, Ft. Lewis College.
"Why Should a Mathematician Worry about Copyright Policy?" by Colleen Quatier, South Dakota School of Mines and Technology.

Wisconsin Section

The fiftieth annual meeting of the Wisconsin Section was held at the University of Wisconsin Center at Fond du Lac, on Friday and Saturday, March 26-27, 1982, with 115 attending. The meeting was held jointly with the meeting of mathematicians from the University of Wisconsin Center System. Highlights included a fiftieth anniversary banquet, presentation of Section T-shirts to 25-year members, and distribution of a sketch of the history of the Section.

Invited Addresses:

"Mathematics Then and Now," by Donald Saari, Northwestern University.
"What in the World is Chaos?" by Donald Saari, Northwestern University.
"Finding Your Way in a Graph," by Ron Graham, Bell Labs.

Contributed Talks:

* "Morality and the Math Class," by Merrill Barneby, University of Wisconsin at LaCrosse.
"Inverting Some Elementary Geometry Theorems," by Orville Bierman, University of Wisconsin at Eau Claire.
* "Complete Traversals of Toroidal Road Maps," by Stephen J. Curran, Beloit College.
"Some Mathematics Unexpectedly Applicable to Computer Graphics and Vice Versa," by Dwight Freund, University of Wisconsin at Waukesha.
"Computer Science and Mathematics--A Look Toward the Wisconsin Section Summer Workshop," by James Freeman, Ripon College.
"Graphics for Mathematics on the Apple Computer," by Ed Gade, University of Wisconsin at Oshkosh.
* "Generalized Fibonacci Numbers by Matrix Methods," by Dan Kalman, University of Wisconsin at Green Bay.
* "Hyperperfect and Unitary Hyperperfect Numbers," by Rudolph Najar, University of Wisconsin at Whitewater.
"Computer Crime," by Tom Napps, Lawrence University.
* "College Remedial Algebra--It's Unreasonable," by Paul A. Taylor, University of Wisconsin at Oshkosh.
"LOGO--A Concrete Approach to Mathematics for Elementary Students," by Don Voils, University of Wisconsin at Oshkosh.
"The High School Connection," by Don Weyers, University of Wisconsin at Superior.

Panel Discussion:

"CUPM Recommendations for a General Mathematical Sciences Program."

Ohio Section

The sixty-sixth annual spring meeting of the Ohio Section was held April 30-May 1, 1982 at Capital University, Columbus, Ohio with 150 individuals in attendance.

Invited Addresses:

"New Initiatives in Pre-College Mathematics Education," by Marcia Sward, Associate Director of MAA.
"Mathematicians of Pre-World War II Germany," by Hans Zassenhaus, Ohio State University.
"Models of Population," by J.D. Faires, Youngstown State University.

Short Presentations:

"Applications of Linear Algebra to Election Studies," by M. Bloxham, Wittenberg University and Essex University, England.
"Fourier Transforms," by B. Ghusayni, University of Toledo.
"The Canterbury Puzzles, or, What Chaucer Missed," by D.J. Horwath, John Carroll University.
"Polyominoes--Pleasurable Pieces for Perception and Perplexing Packing Puzzles," by R.G. Laatsch, Miami University.
"For Rubik's Yet to Come," by W.A. McWorter, Jr., Ohio State University.
"The Census Taker Problem," by L.F. Meyers, Ohio State University.
"Countable Perfect Metric Spaces," by G.V. Ramesh, University of Toledo.
"The Volume of a Cone," by J.H. Riley, Jr., Ohio Northern University.
"Report on the Mathematical Content of the 1982 American High School Mathematics Examination," by L.J. Schneider, John Carroll University.

"Accuracy of Eigenvalues of Symmetric Matrices," by K.A. Schueller, Youngstown State University.

At a special session, 22 student papers were presented.

Oklahoma-Arkansas Section

This meeting was held at the University of Arkansas at Fayetteville on March 26-27, 1982 in Fayetteville, Arkansas. Approximately 125 were in attendance.

Invited Addresses:

"The Isoperimetric Theorem," by Ivan Niven, University of Oregon.
"Does Mathematics Have Elements?," by Paul R. Halmos, Indiana University.

Short Presentations:

"The Algebraic Structure of the Rubik's Cube," by Benny Evans, Oklahoma State University.
"The Geometric Structure of a Real Number," by T. Sekaguchi, University of Arkansas.
"When is a Detour Cheaper?" by Philip Almes, East Central University.
"A Nonlinear Least Squares Problem," by L. Franklin Kemp, Amoco Production Research.
"Exponential Stability Equals Stability to Linearization," by Sherwin Skar, Oklahoma State University.
"Oscillatory Solutions of the Differential Equation $Lny + py = 0$," by Bennette Harris and Marvin Keener, Oklahoma State University.
"A Survey of Norm Attaining Operators on Banach Spaces," by Jerry Johnson, Oklahoma State University.
"Algebras of Bounded Sequences Induced by $H^{\wedge}$," by Robert C. Smith, University of Arkansas.
"Some Comment on Caley's Theorem," by Elouise Shorter, University of Arkansas at Pine Bluff.
"Questions About Queens: The N-Queens Problem," by Ken Parker, Oklahoma Christian College.
"Program HIPLOT: High Resolution Plotting on the Commodore Printer," by Tim Best, Hendrix College.
"Totally Real Rooted Omnimonics," by Karim Keith Carter, University of Arkansas at Pine Bluff.
"How to Recognize the Abelian Subgroups of the Group M of Mobius Transformations," by Sheila P. Harris, University of Arkansas at Pine Bluff.
"Taxicab Relativity," by Benjamin Schumacher, Hendrix College.
"Optimal Satellite Search Techniques," by Jeff Thompson, East Central University.
"Computer Simulation of a Markov Chain," by Jerry Coker, Hendrix College.
"Computeronics: A Cost Effective Project for Accelerated Middle School Mathematics Students," by Cecil McDermott, Hendrix College.
"Teaching Techniques Gleaned from the Babylonians," by Robert D. McMillan, Oklahoma Christian College. "The Secondary Mathematics Teacher Shortage in Arkansas," by William Orton, University of Arkansas.
"Applied Math Seminars, Another Approach," by Verbal Snook, Oral Roberts University.
"The MAWIS Project," by Glenda Owens and John Watson, Central State University and Arkansas Technology University.
"S-Automata with Some Examples," by Harold Davenport, University of Arkansas at Little Rock.
"Transformations Belonging to Inverse Semigroups," by Boris Schein, University of Arkansas.
"A Simple Proof of the Noncommutative Eakin Theorem," by Richard Resco, University of Oklahoma.
"Amalgamations and Divisible Embeddings for Lattice Ordered Groups," by Wayne Powell, Oklahoma State University.
"Characterizations of Functionally Compact Spaces," by Larry Herrington and Paul Long, Louisiana State University of Alexandria.
"The Closed Graph and Minimal Topological Spaces," by Ray Hamlett, East Central University.
"Para-Continuous Functions," by Paul Long and Larry Herrington, University of Arkansas.
"The Picard Group of a Compact Complex Nil-Manifold," by Robert Fisher, University of Oklahoma.
"Finite Subdirect Products of Rings," by Joel Haack, Oklahoma State University.
"The Jacobson-Relational Radical and Structure of a Semiring," by Lide Li, University of Arkansas.
"Properties of Congruences on Regular Semigroups," by Bernard Madison, University of Arkansas.
"The Conjugacy Problem for HNN Extensions of Finitely Generated Abelian Groups," by Farrokh Abedi, Oklahoma State University.
"A Categorical Proof of a Theorem of Jonssón and Tarski," by Zeev Barel, Hendrix College.
"The Evaluation of Travel Time: An Example in the Application of Queueing Theory," by L. Wade Nixon, Hendrix College.
"Properties of Functions of Bounded Variation," by Darryl Linde, University of Arkansas.
"Mathematical Modeling and the Wine Cellar," by Dennis Meredith, Hendrix College.
"Lines in a Metric Space," by Carol Smith, Hendrix College.
"An Application of Green's Theorem," by Susan Mengal, Central State University.
"Airspace Requirements of Holding Patterns," by Leslie Ribera, Central State University.
"Mathematical Modeling," by Sara Churchill, Oral Roberts University.
"Modeling the Four-Five Hole Problem," by Daryl Ezzo, Oral Roberts University.
"The Microwave Reflection Problem," by Kathy Davenport and Bob Bailey, Arkansas Technology University.
"Using the Apple II Plus in the Classroom," by Jim Peden, Arkansas Technology University.

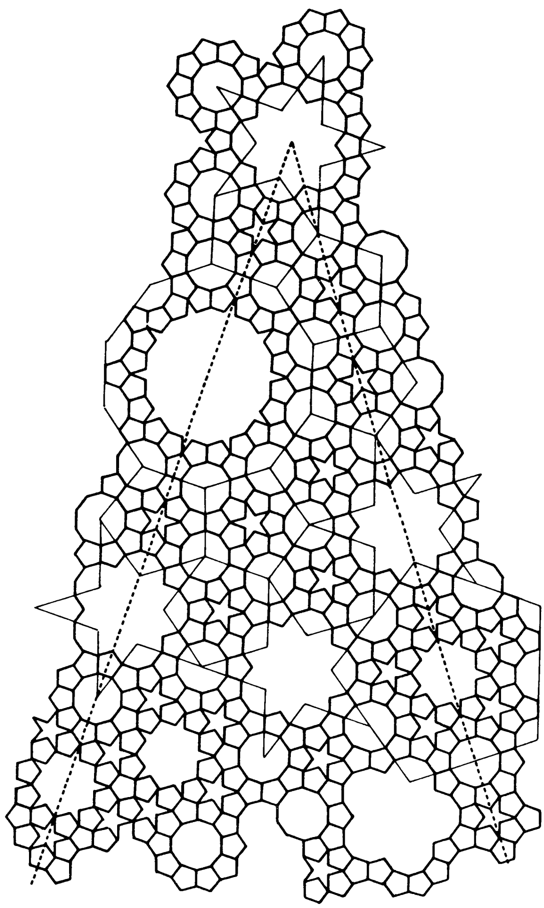


FIG. 6. A patch with $\rho \approx 8.6$ obtained by "decomposition" from the tiling in Fig. 5 and addition of some tiles around the boundary.

8. I. El-Said and A. Parman, Geometric Concepts in Islamic Art, World of Islam Festival Publishing Co., London, 1976.

9. L. Fejes Tóth, Reguläre Figuren, Akadémiai Kiadó, Budapest, 1965. English translation: Regular Figures, Pergamon, New York, 1964.

10. B. Grünbaum and G. C. Shephard, Tilings and Patterns, Freeman, San Francisco, 1982.

11. J. Kepler, Harmonice Mundi. Lincii, 1619. Reprints: Culture et Civilisation, Bruxelles, 1968; Johannes Kepler Gesammelte Werke, M. Caspar, editor, Band VI, Beck, München, 1940. Annotated German translation: M. Caspar, Welt-Harmonik, Oldenbourg, München, 1939. An annotated English translation of Book II appears in: J. V. Field, Kepler's star polyhedra, Vistas in Astronomy, 23(1979) 109–141.

12. D. Schattschneider, Tiling the plane with congruent pentagons, Math. Mag., 51(1978) 29–44.

13. ______, In praise of amateurs, The Mathematical Gardner, edited by D. A. Klarner, Prindle, Weber & Schmidt, Boston, 1981, pp. 140–166 + 5 plates in color.

14. D. Wade, Pattern in Islamic Art, Overlook Press, Woodstock, NY, 1976.

MISCELLANEA

83. Mathematics, like Dialectics, is an organ of the inner, higher mind; in practice, it is an art like eloquence. In both, nothing counts but the form; the content is irrelevant.

—J. W. v. Goethe, *Maximen und Reflexionen*, no. 605.

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

SYLVESTER'S THEOREM FOR MATRICES WITH SMOOTH ENTRIES

GREGORY A. FREDRICKS
Department of Mathematics, Texas Tech University, Lubbock, TX 79409

We present here a classification theorem for matrices with smooth entries. Our proof uses ideas from differential geometry and the theory of fibre bundles.

Let $S(n)$ denote the set of symmetric $n \times n$ matrices and $GL(n)$ denote the set of nonsingular $n \times n$ matrices, both with real entries. Recall that $A, B \in S(n)$ are *conjugate* if there exists $G \in GL(n)$ such that

$$GA^tG = B.$$

A well-known classical result, sometimes called Sylvester's Theorem (see Perlis [**2**, p. 91]), states that A and B are conjugate if and only if they have the same number of positive and negative eigenvalues. Equivalently A and B are conjugate if and only if they belong to the same orbit of the group action of $GL(n)$ on $S(n)$ defined by

$$\nu(G, A) \equiv GA^tG.$$

In the orbit consisting of the symmetric matrices with p positive and q negative eigenvalues (p, q nonnegative integers with $p + q \leqslant n$) we have the matrix

$$D_{p,q} \equiv \operatorname{diag}(1,\dots,1,-1,\dots,-1,0,\dots,0),$$

with p 1's and q -1's. Thus we have the following version of

SYLVESTER'S THEOREM. *Given $A \in S(n)$, there exists $G \in GL(n)$ such that*

$$GA^tG = D_{p,q}$$

if and only if A has p positive and q negative eigenvalues.

The corresponding result for matrices with smooth, i.e., infinitely differentiable, entries is

THEOREM. *Let U be a smoothly contractible open set in $\mathbb{R}^k$ and let $A\colon U \to S(n)$ be a smooth map. There exists a smooth map $G\colon U \to GL(n)$ such that*

$$GA^tG = D_{p,q} \quad \text{on } U$$

if and only if A has p positive and q negative eigenvalues at each point of U.

This theorem can be considered as a result in the theory of functional equations and was announced in [**3**, p. 37]. An application of the theorem to the geometric theory of second order linear partial differential operators is given in [**1**].

Proof of Theorem. The necessity of the eigenvalue condition is clear from the classical Sylvester's Theorem.

To show that the eigenvalue condition is sufficient, we fix p and q, define $r \equiv p + q$, and let D denote $D_{p,q}$. Now let $S(n, r)$ denote the submanifold of $S(n)$ consisting of all symmetric $n \times n$ matrices of rank r and let $F\colon U \times GL(n) \to S(n, r)$ be the smooth map defined by

$$F(x, g) \equiv gA(x)^tg.$$

Since the components of F in the usual coordinates are given by

$$F\left(x,\left(g_j^i\right)\right)_{(l,m)}=\sum_{i,j}a_{ij}(x)g_i^lg_j^m,$$

one can easily show that

$$\operatorname{rank}\frac{\partial(F)}{\partial(g)}(x,g)=n+(n-1)+\cdots+(n-r+1)$$

$$=\dim S(n,r)\quad \textit{for each } (x,g)\in F^{-1}(D).$$

(The implicit function theorem now implies the existence of G locally.) We therefore conclude that D is a regular value of F and hence, by Spivak [**4**, p. 2-30], that

$$E\equiv F^{-1}(D)=\left\{(x,g)\in U\times GL(n)\mid gA(x)\,{}^tg=D\right\}$$

is a submanifold of $U\times GL(n)$. It follows from the eigenvalue condition and the classical Sylvester's Theorem that the obvious map π from E to U is a projection *onto* U and hence a map of constant maximal rank. Moreover, if

$$Z\equiv\left\{g\in GL(n)\mid gD\,{}^tg=D\right\},$$

then $\pi^{-1}(x)=\{(x,Zg)\}$, where (x,g) is any element of $\pi^{-1}(x)$. Since $GL(n)$ is a Lie group, we see that the triple (E,π,U) is a smooth fibre bundle (with fibre Z and group Z). Note that the desired map $G\colon U\to GL(n)$ is any smooth section of this bundle. Since U is smoothly contractible, we see by Steenrod [**5**, p. 53], that (E,π,U) is trivial and hence has smooth sections.

Note from the proof that this Sylvester's Theorem holds, more generally, for matrices with entries which are smooth functions on a manifold with vanishing first cohomology group.

References

1. G. A. Fredricks, The geometry of second order linear partial differential operators, to appear.

2. S. Perlis, Theory of Matrices, Addison-Wesley, Cambridge, MA, 1952.

3. Proceedings of the Eighteenth International Symposium on Functional Equations, Centre for Information Theory, Univ. of Waterloo, Waterloo, Ontario, Canada, 1980.

4. M. Spivak, A Comprehensive Introduction to Differential Geometry, vol. 1, Publish or Perish, Boston, MA, 1970.

5. N. E. Steenrod, The Topology of Fibre Bundles, Princeton Univ. Press, Princeton, NJ, 1951.

INJECTIVE HOLOMORPHIC MAPPINGS

JEAN-PIERRE ROSAY

Mathématiques, Université de Provence, Pl. V. Hugo, 13331 Marseille Cedex 3, France

My aim is to give a very simple (and easy to remember) proof of the following well-known fact:

PROPOSITION. *Let Ω be an open set in $\mathbb{C}^n$ and F a holomorphic mapping from Ω to $\mathbb{C}^n$. If F is one-to-one, then the Jacobian determinant of F does not vanish (and therefore F is biholomorphic onto* $F(\Omega)$).

My only excuse for writing this note is that I found rather surprisingly complicated the proofs that I read in the literature: [**1**, Theorem 7, p. 179], [**2**, Theorem 7.3, p. 112 (consequence of Osgood's Theorem 7.2, p. 105)], [**3**, Theorem 5, p. 86], [**4**, Theorem 15.1.8, p. 302].

We write $F=(f_1,\dots,f_n)$. Let (JF) denote the Jacobian determinant of F; (JF) $=\det(\partial f_i/\partial z_j)_{1\leq i,j\leq n}$.

The result we want to prove is clear if $n=1$; we then proceed by induction and assume that it

is already proved for the dimension $n-1$. I could almost stop here!

CLAIM. *If U is a neighborhood of 0 in $\mathbb{C}^n$, F: $U \to \mathbb{C}^n$ is holomorphic and one-to-one, and $(\partial f_i/\partial z_j)(0) \neq 0$ for some i and j, then* (JF) (0) $\neq 0$.

Proof of the Claim. Without loss of generality we can assume that $F(0) = 0$ and $(\partial f_1/\partial z_1)(0) = a \neq 0$. Define $H(z) = (f_1(z), z_2, \ldots, z_n)$. Since (JH)(0) $= a \neq 0$, H is a biholomorphic map of a neighborhood of 0 onto a neighborhood V of 0. Therefore

$$G(w) = F(H^{-1}(w))$$

defines a holomorphic one-to-one map $G = (g_1, \ldots, g_n)$ of V into $\mathbb{C}^n$. For $w = (w_1, \ldots, w_n)$ in V, we see that

$$g_1(w) = f_1(H^{-1}(w)) = w_1$$

and (JF)(0) $= a$(JG)(0). Setting $w' = (w_2, \ldots, w_n)$, define

$$\tilde{G}(w') = (g_2(0, w'), \ldots, g_n(0, w')).$$

Then $\tilde{G}$ is holomorphic and one-to-one in some neighborhood of $0'$ in $\mathbb{C}^{n-1}$, so that $(J\tilde{G})(0') \neq 0$ by induction hypothesis. But $(J\tilde{G})(0') = (JG)(0)$. Hence (JF)(0) $\neq 0$, as desired.

Proof of the Proposition. If not empty, the set defined by (JF) $= 0$ is an analytic variety of complex codimension 1. On this variety, by the claim, the partial derivatives of all the functions f_i are identically 0. Therefore F is locally constant on this variety (this fact is really elementary, at least at the regular points of the variety), which of course violates the hypothesis that F is one-to-one.

REMARK. We can even avoid any consideration of analytic varieties, giving another proof.

Second proof of the Proposition. We can assume that Ω is bounded, $0 \in \Omega$, and $F(0) = 0$. We want to prove that (JF)(0) $\neq 0$. Let g be the function defined by $g(z) = \sum_{i=1}^{n} |f_i(z)|^2$ and if necessary replace Ω by a slightly smaller set so that g is bounded away from 0 near $\partial\Omega$. Choose $\epsilon > 0$ small enough so that the closure of $\{z \in \Omega\colon g(z) < \epsilon\}$ is contained in Ω. By Sard's Theorem, there exists $\alpha \in (0, \epsilon)$ such that grad $g \neq 0$ on $g^{-1}(\alpha)$. Let $\mathcal{O} = \{z \in \Omega\colon g(z) < \alpha\}$. The claim implies that JF $\neq 0$ on $\partial\mathcal{O}$. Thus JF $\neq 0$ on $\mathcal{O}$, because for $n > 1$ the zeroes of a holomorphic function propagate out to the boundary (a basic and easy fact). In particular, (JF)(0) $\neq 0$, as desired.

References

1. S. Bochner, W. Martin, Functions of Several Complex Variables, Princeton Univ. Press, Princeton, N.J., 1948.
2.. B. Fuks, Introduction to the theory of analytic functions in several complex variables, Translations of Math. Monographs, v. 8, Amer. Math. Soc., 1963.
3. R. Narasimhan, Several complex variables, Chicago Lect. in Math., 1971.
4. W. Rudin, Function theory in the unit ball of $\mathbb{C}^n$, Grund. der Math. W. 241, Springer Verlag, 1980.
5. (Added in proof) S. Krantz, Function Theory of Several Complex Variables, Wiley, 1982, p. 397.

ANSWER TO "PHOTO" ON PAGE 567

M. H. Stone's name is visible in Stone's theorem (about unitary groups), in the Stone representation theorem (about Boolean algebras), in the Stone-Čech compactification, and in the Stone-Weierstrass theorem. The photo was taken in March, 1982.

PROBLEMS AND SOLUTIONS

EDITED BY DAVID BORWEIN, J. L. BRENNER, AND VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

The Sum $\Sigma \pm 2^i$, with m or Fewer Terms

S 32 [1980, 487]. *Proposed by Walter Feit, Yale University.*

For integers m, n with $m > 0$, let $N_m(n)$ denote the number of ways that a subset S of $\{0, 1, \ldots, m-1\}$ and a sequence of $\pm$ signs can be chosen so that $n = \Sigma_{i\in S} \pm 2^i$. (The empty sum is 0 by definition.) For instance, $N_m(\pm[2^m - (-1)^m]) = 1 + (-1)^m$; and so $\Sigma N_m(k) = 2 + (-1)^m$, this sum being taken over those values of k for which $[2^m - (-1)^m] \mid k$. Find $\Sigma N_m(k)$, if the sum is now taken over those values of k for which $\{[2^m - (-1)^m]/3\} \mid k$.

Solution by the proposer. The answer is $2 + (-1)^m + 2L_m$, where

$$L_m = \left(\frac{1+\sqrt{5}}{2}\right)^m + \left(\frac{1-\sqrt{5}}{2}\right)^m$$

is the mth Lucas number.

Proof. Let $\alpha_m = (2^m - (-1)^m/3)$ and let $a_m = N_m(\alpha_m) + N_m(2\alpha_m)$. Since $N_m(n) = N_m(-n)$ it suffices to show that $a_m = L_m$.

By inspection $a_1 = L_1 = 1$, $a_2 = L_2 = 3$ and $a_3 = L_3 = 4$.

If $2\alpha_m$ is of the form $\Sigma \pm 2^i$, then the term 2^{m-1} must occur as

$$\sum_{i=0}^{m-2} 2^i = 2^{m-1} - 1 < 2\alpha_m.$$

As $2\alpha_m - 2^{m-1} = \alpha_{m-1} + (-1)^{m-1}$ it follows that $N_m(2\alpha_m) = N_{m-1}(\alpha_{m-1} + (-1)^{m-1})$. Thus

$$a_m = N_m(\alpha_m) + N_{m-1}\left(\alpha_{m-1} + (-1)^{m-1}\right).$$

Let $\beta_m = \alpha_m$ or $\alpha_m + (-1)^m$. If β_m is of the form $\Sigma \pm 2^i$ and $m \geqslant 2$, then one of the terms 2^{m-1} or 2^{m-2} must occur as

$$\sum_{i=0}^{m-3} 2^i = 2^{m-2} - 1 < \beta_m.$$

Since

$$\alpha_m - 2^{m-1} = -\alpha_{m-1} \quad \text{and} \quad \alpha_m - 2^{m-2} = \alpha_{m-2}$$

it follows that

$$N_m(\alpha_m) = N_{m-1}(\alpha_{m-1}) + N_{m-2}(\alpha_{m-2})$$
$$N_m\big(\alpha_m + (-1)^m\big) = N_{m-1}\big(\alpha_{m-1} + (-1)^{m-1}\big) + N_{m-2}\big(\alpha_{m-2} + (-1)^{m-2}\big).$$

Therefore $a_{m+2} = a_{m+1} + a_m$ for $m \geqslant 2$. This implies the result as $L_{m+2} = L_{m+1} + L_m$ for $m \geqslant 1$.

REMARK. The argument also shows that

$$N_m(\alpha_m) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{m+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{m+1}\right],$$

and

$$N_m(2\alpha_m) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{m-1} - \left(\frac{1-\sqrt{5}}{2}\right)^{m-1}\right].$$

An Application of Multigrades to Banach Spaces of Sequences

S 33 [1980, 487]. *Proposed by Simeon Reich, University of Southern California.*

Let $\{x_n\}$, $n = 1,2,\ldots$, be a sequence in L^p, $1 < p < \infty$. Suppose that for some y in L^p, $\lim_{n\to\infty}\|x_n - ty\|$ exists for all $0 \leqslant t \leqslant 1$. For which values of p does it follow that $\lim_{n\to\infty}\|x_n - ty\|$ exists for *all* $t \geqslant 0$?

What is the situation in other Banach spaces?

For $p = 2n$ (an even positive integer), the conclusion is valid. See Ronald E. Bruck and Simeon Reich, *A general convergence principle in nonlinear functional analysis*, Nonlinear Analysis, Theory, Methods & Applications 4 (1980) 939–950. (For these values of p, the norm is polynomial.) The assertion may be false for all other values of p.

Bruce Reznick (University of Illinois) kindly supplied a class of counterexamples. Given $n > 0$, there exist m and positive numbers a_i, b_i, $1 \leqslant i \leqslant m$, with $\Sigma_{a_i}^{2k} = \Sigma_{b_i}^{2k}$ for all $1 \leqslant k \leqslant n$, but $\Sigma_{a_i}^{2n+1} \neq \Sigma_{b_i}^{2n+1}$. (The a_i, b_i can even be positive integers, but that is not essential here.)

Note that

$$\sum(t+a_i)^{2n+1} + \sum(t-a_i)^{2n+1} = 2\sum\binom{2n+1}{2k}t^{2n+1-2k}a_i^{2k}$$
$$= \sum(t+b_i)^{2n+1} + \sum(t-b_i)^{2n+1}.$$

We take $\mathbf{z} = (z_1,\ldots,z_s) \in l^{2n+1}(\mathbb{R}^s)$, and then $\|z\|^{2n+1} = \Sigma|z|^{2n+1}$. Set

$$\mathbf{u} = (a_1, -a_1, a_2, -a_2,\ldots,a_m,-a_m), \qquad \mathbf{v} = (b_1,-b_1,b_2,-b_2,\ldots,b_m,-b_m),$$
$$\mathbf{1} = (1,1,\ldots,1).$$

Take $T \geqslant \max\{|a_i|, |b_i|\}$. Then for $t \geqslant T$, we find

$$\|\mathbf{u} + t\mathbf{1}\|^{2n+1} = \sum(t+a_i)^{2n+1} + (t-a_i)^{2n+1} = \|\mathbf{v} + t\mathbf{1}\|^{2n+1}.$$

In particular, with $\mathbf{x} = \mathbf{u} + (T+1)\mathbf{1}$, $\mathbf{y} = \mathbf{v} + (T+1)\mathbf{1}$, it is true that $\mathbf{x} \neq \mathbf{y}$, $\|\mathbf{x} - t\mathbf{1}\| = \|\mathbf{y} - t\mathbf{1}\|$ for $0 \leqslant t \leqslant 1$, but not for $t = T+1$.

For $2n+1 = 3$, take $\{a_i\} = \{1,3\}$, $\{b_i\} = \{2,\sqrt{6}\}$, $T = 3$ to get a numerical counterexample.

Area of a Pentagon in Terms of Areas of Inscribed Triangles

S 34 [1980, 574]. *Proposed by O. Bottema, Delft, The Netherlands.*

In a plane, a non-self-intersecting pentagon $A_1A_2A_3A_4A_5$ is given. No three of the vertices A_i

are collinear and (ijk) denotes the (signed) area of the (oriented) triangle $A_iA_jA_k$. Furthermore

$$(124) = a_1, \quad (235) = a_2, \quad (341) = a_3, \quad (452) = a_4, \quad (513) = a_5.$$

Determine the area of the pentagon $A_1A_2A_3A_4A_5$.

[The analogous problem, with (123), (234), (345), (451), and (512) being given, was solved by Gauss in 1823. The Canadian journal *Eureka* (vol. 3, no. 8, 1977, p. 240) gave a reproduction of his solution.]

Solution by Vojislav Petrović, University of Novi Sad, Yugoslavia. Denote by $(12345) = S$, $(123) = \alpha_1$, $(234) = \alpha_2$, $(345) = \alpha_3$, $(451) = \alpha_4$, $(512) = \alpha_5$ (signed) areas of (oriented) pentagon $A_1A_2A_3A_4A_5$ and triangles $A_1A_2A_3$, $A_2A_3A_4$, $A_3A_4A_5$, $A_4A_5A_1$, $A_5A_1A_2$. From the identity due to Möbius:

$$(123)(145) + (142)(135) + (125)(134) = 0$$

we have

$$\alpha_1\alpha_4 - \alpha_1\alpha_5 + \alpha_5\alpha_3 = 0. \tag{1}$$

Solving the system of equations

$$\begin{aligned} \alpha_1 + \alpha_3 &= S - a_5 & \alpha_3 + \alpha_5 &= S - a_2 \\ \alpha_2 + \alpha_4 &= S - a_1 & \alpha_4 + \alpha_1 &= S - a_3 \\ & & \alpha_5 + \alpha_2 &= S - a_4 \end{aligned}$$

we have

$$\begin{aligned} \alpha_1 &= (S + a_1 + a_2 - a_3 - a_4 - a_5)/2 \\ \alpha_4 &= (S - a_1 - a_2 - a_3 + a_4 + a_5)/2 \\ \alpha_5 &= (S + a_1 - a_2 - a_3 - a_4 + a_5)/2. \end{aligned} \tag{2}$$

Substituting (2) in (1) we get

$$S = (a_1 + a_2 + a_3 + a_4 + a_5)^2 - 4(a_1a_3 + a_2a_4 + a_3a_5 + a_4a_1 + a_5a_2)^{1/2}.$$

Essentially the same solution was given by Kit Hanes, Eastern Washington University; M. Hausner, New York University; L. Kuipers, Mollens, Vs, Switzerland; O. P. Lossers, Eindhoven Institute of Technology, Netherlands; Leroy F. Meyers, Ohio State University, and the proposer.

A short proof of the Möbius identity was furnished by some of the solvers. O. P. Lossers' proof is as follows:

Putting $A_1A_k = \mathbf{v}_k$ $(k = 2, 3, 4, 5)$ and using the identity

$$(\mathbf{v}_2 \times \mathbf{v}_3, \mathbf{v}_4 \times \mathbf{v}_5) + (\mathbf{v}_2 \times \mathbf{v}_4, \mathbf{v}_5 \times \mathbf{v}_3) + (\mathbf{v}_2 \times \mathbf{v}_3, \mathbf{v}_3 \times \mathbf{v}_4) = 0,$$

we obtain: $b_1b_4 - (124) \cdot (135) + b_5(134) = 0$, or:

$$b_1b_4 - (x - b_2 - b_4)(x - b_1 - b_3) + b_5(x - b_1 - b_4) = 0,$$

that is: $x^2 - bx + c = 0$.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by February 28, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2962. *Proposed by M. S. Klamkin, University of Alberta, Canada.*

It is known that if the circumradii R of the four faces of a tetrahedron are congruent, then the four faces of the tetrahedron are mutually congruent (i.e., the tetrahedron is isosceles) [**1**]. It is also

known that if the inradii r of the four faces of a tetrahedron are congruent, then the tetrahedron need not be isosceles [2]. Show that if Rr is the same for each face of a tetrahedron, the tetrahedron is isosceles.

References

1. Crux Mathematicorum, 6 (1980) 219.
2. Crux Mathematicorum, 4 (1978) 263.

E 2963. *Proposed by Calin P. Popescu, student, Bucharest, Rumania.*

Let $A_1A_2A_3$, $A_1'A_2'A_3'$ be two equilateral triangles in the plane. Construct circles γ_i $[\gamma_i']$ with radii r_i $[r_i']$ and centers A_i $[A_i']$, $i = 1,2,3$ $[i = 1,2,3]$. Suppose further that r_i $[r_i']$ are geometric progressions with ratio a positive integer. When can the six circles be concurrent?

E 2964. *Proposed by M. Slater, University of Bristol, U.K.*

Suppose f is a real function having a left-hand derivative f_L' and a right-hand derivative f_R' at every point of the real interval I. We say that real numbers a, b *have opposite sign* provided $ab \leqslant 0$. Show that the following are equivalent:

(A) *One* of the functions f_L', f_R' takes values of opposite sign at some *two* points p, q of I.

(B) The *two* functions f_L', f_R' take values of opposite sign at some *one* point r of I.

E 2965. *Proposed by Bu Qi-yue, Shanghai Jiao-tung University, People's Republic of China.*

Let $0 < h$, $A \subseteq [a, b]$, A measurable. Prove

$$\frac{1}{2h}\int_a^b m[A \cap (x-h, x+h)]\,dx \leqslant mA.$$

E 2966. *Proposed by P. J. Giblin, University of Liverpool, U.K.*

A, B, P_1, P_2, P_3 are distinct points in the plane. P_iP_jA, P_iP_jB are proper triangles, i.e., no two of P_1, P_2, P_3 are collinear with A or with B. The anticlockwise angles from AP_1 to AP_2, AP_1 to AP_3, BP_1 to BP_2, BP_1 to BP_3 are θ_1, θ_2, ϕ_1, ϕ_2. If $a_i = AP_i$, $b_i = BP_i$ and if the relations

$$\frac{\sin\theta_1}{\sin\phi_1}\frac{a_3}{b_3} = \frac{\sin\theta_2}{\sin\phi_2}\frac{a_2}{b_2} = \frac{\sin(\theta_2-\theta_1)}{\sin(\phi_2-\phi_1)}\frac{a_1}{b_1}$$

hold, show that the angle AP_iB has the same pair of bisectors as one of the angles of the triangle $P_1P_2P_3$. (Possibly the internal bisector of one angle is the external bisector of the other.)

E 2967. *Proposed by Jordi Dou, Barcelona, Spain.*

Divide a circle into four equiareal parts with (i) arcs (ii) line segments of minimal total length.

SOLUTIONS OF ELEMENTARY PROBLEMS

Six Segments Defined by a Point Interior to a Triangle

E 2716* [1978, 384]. *Proposed by Jack Garfunkel, Flushing, NY.*

Let ABC be a triangle with P an interior point. Let A', B', C' be the points where the perpendiculars drawn from P meet the sides of ABC. Let A'', B'', C'' be the points where the lines joining P to A, B, C meet the corresponding sides of ABC.

Prove or disprove that $A'B' + B'C' + C'A' \leqslant A''B'' + B''C'' + C''A''$.

Solution by C. S. Gardner, Austin, Texas.

THEOREM. The proposed inequality is correct. A proof follows.

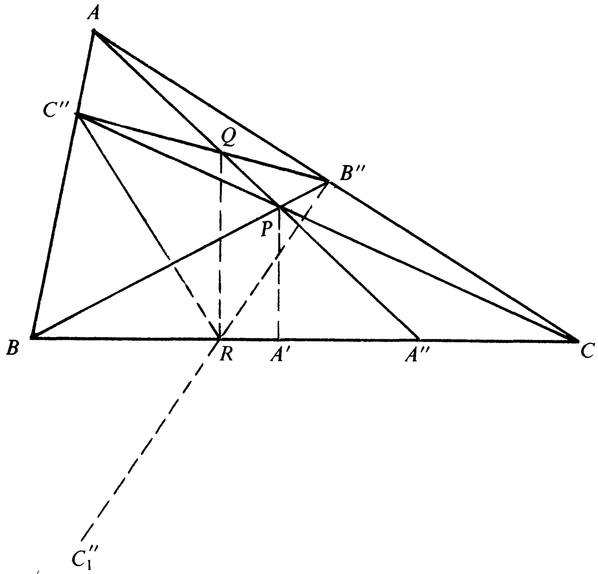


FIG. 1.

LEMMA 1. See Figure 1. $A'B'' + A'C'' \leqslant A''B'' + A''C''$ (and the two similar inequalities).

Let $B''C''$ meet AP in Q and let R be the foot of the perpendicular from Q on BC. By the harmonic properties of the complete quadrilateral $AB''PC''$ it is seen that RQ, RC separate RC'', RB'' harmonically. Being perpendicular, RQ and RC accordingly bisect $B''RC''$ internally and externally. If C_1'' is the image of C'' on reflection in BC, the path $B''RC_1''$ is straight. Since A' is between R and A'',

$$A'B'' + A'C'' = A'B'' + A'C_1'' \leqslant A''B'' + A''C_1'' = A''B'' + A''C''. \qquad \square$$

From here on we assume that the vertices of the triangle are so labelled that the point P is in the closed triangle OA_0C, where O is the orthocenter of ABC and A_0 is the foot of the perpendicular from A on BC. Note if one vertex angle is not acute, it is the angle at A, and P is in the triangle AA_0C. The points A'', B'', C'' are on the segments $A'C$, $B'C$, $C'B$ respectively. The points A', B' are interior to the sides BC, CA. The point C' is either interior to BA, or—and only if A is obtuse—on BA produced.

Let the angles $PA'B'$, $PA'C'$, $PB'A'$, $PB'C'$, $PC'A'$, $PC'B'$ be denoted by $a_2, a_3, b_1, b_3, c_1, c_2$ respectively. If C' is interior to BA, all these angles are acute. If C' is on BA produced, then b_3 is obtuse and all the other angles are acute. We note $b_1 + b_3 < A'PC' = \pi - B < \pi$, whence $b_1 < \pi - b_3$.

LEMMA 2. See Figure 3. $b_1 \leqslant b_3$ and $\sin b_1 \leqslant \sin b_3$.

Let CP meet AA_0 in P_1, and let the feet of the perpendiculars from P_1 on AC, AB be B_1', C_1'. Then $b_1 = P_1B_1'A_0$, $b_3 \geqslant P_1B_1'C_1'$. The quadrilaterals $P_1B_1'CA_0$, $P_1B_1'AC_1'$ are cyclic, since the angles at B_1', A_0, C_1' are right angles. Hence

$$b_1 = \pi/2 - A_0P_1C \leqslant \pi/2 - C_1'P_1A \leqslant b_3.$$

(The angles $P_1B_1'A_0$, P_1CA_0 are equal, since the points P_1, B_1', C, A_0 are on a circle; and likewise $P_1AC_1' = P_1B_1'C_1'$.) If b_3 is acute it follows immediately that $\sin b_1 \leqslant \sin b_3$. If b_3 is obtuse, then $\pi - b_3$ is acute, and then since $b_1 < \pi - b_3$, it again follows that $\sin b_1 \leqslant \sin b_3$. $\square$

LEMMA 3. See Figure 2. At least one of the inequalities $c_1 \leqslant c_2$, $a_2 \leqslant a_3$ holds.

We have $PA'\sin a_2 = PB'\sin b_1$, $PB'\sin b_3 = PC'\sin c_2$ and $PC'\sin c_1 = PA'\sin a_3$. Multi-

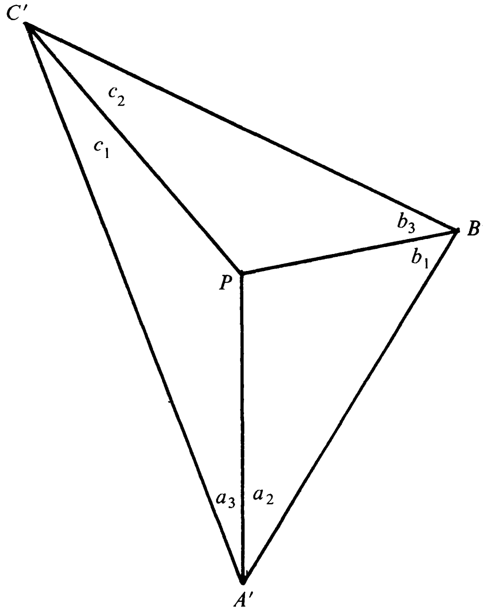


FIG. 2.

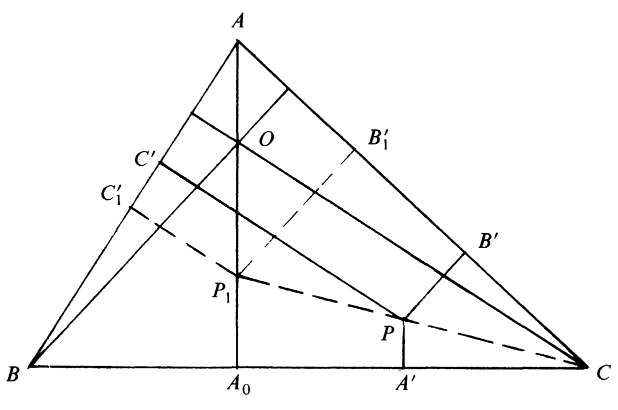


FIG. 3

plying these together, we get

$$\sin a_2 \sin b_3 \sin c_1 = \sin a_3 \sin b_1 \sin c_2.$$

Since $\sin b_1 \leqslant \sin b_3$, it follows that either $\sin c_1 \leqslant \sin c_2$ or $\sin a_2 \leqslant \sin a_3$. Since c_2 and a_3 are acute, it follows that either $c_1 \leqslant c_2$ or $a_2 \leqslant a_3$. □

THEOREM, Case 1a. See Figure 4. Suppose $c_1 = PC'A' \leqslant c_2 = PC'B'$ and that C' is interior to BA.

Let $C'D$ be the reflection of the ray $C'B'$ in the line $C'P$. (Meaning: the angle $DC'P =$ the angle $B'C'P$.) Since $c_1 \leqslant c_2$ and c_2 is acute, the ray $C'D$ is inside the angle $A'C'C''$ and hence meets $A'C''$ at a point D between A' and C''. Likewise, since $b_1 \leqslant b_3$ and b_3 is acute, the reflection

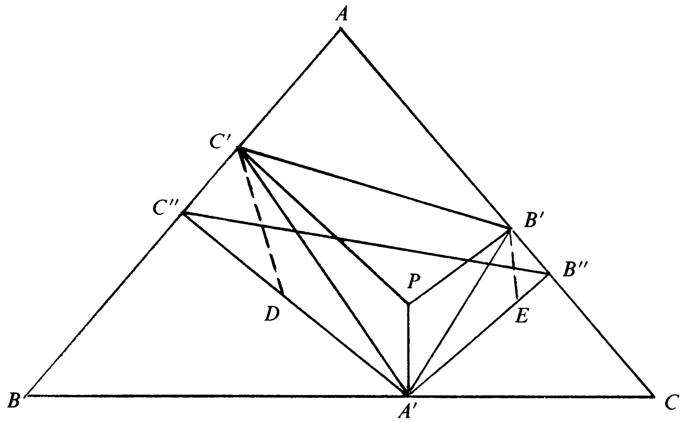


FIG. 4

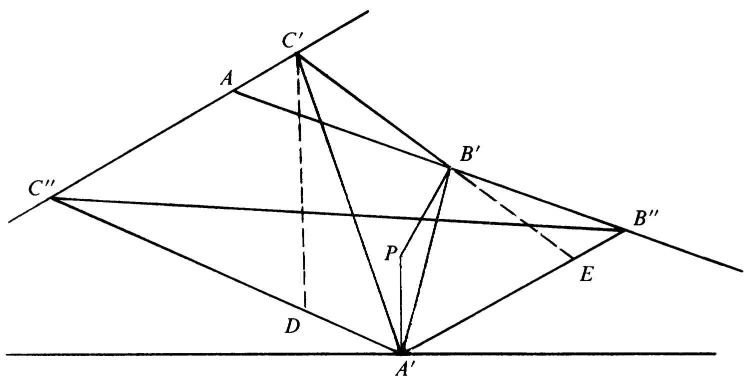


FIG. 5

$B'E$ of $B'C'$ in $B'P$ meets $A'B''$ at a point E between A' and B''. The path $DC'B'E$ is the shortest path from D to E meeting AB, AC in that order.

Proof. If D_1 is the reflection of D in AB and E_1 is the reflection of E in AC, then the path $D_1CB'E_1$ is straight. So $DC'B'E \leqslant DC''B''E$, and, from Lemma 1,

$$\begin{aligned} A''C'' + C''B'' + B''A'' &\geqslant A'D + DC''B''E + EA' \\ &\geqslant A'D + DC' + C'B' + B'E + EA' \\ &\geqslant A'C' + C'B' + B'A'. \end{aligned}$$ □

THEOREM, Case 1b. See Figure 5. Suppose $c_1 \leqslant c_2$, as before, but C' is on BA produced.

Construct the point D exactly as before. However, E is constructed by producing the line $C'B'$; it meets $A'B''$ at E between A' and B''. This follows from the two facts that $b_3 - \pi/2 \leqslant \pi/2 - b_1$ and that b_3 is obtuse. After this modification, the rest of the proof is the same as in Case 1a. □

THEOREM, Case 2a. See Figure 6. Suppose $a_2 = PA'B' \leqslant a_3 = PA'C'$, and that C' is interior to BA.

Reflect $A'B'$ in $A'P$. The reflected ray can be shown to meet $A''C'$ at a point D between A'' and C'. Likewise the reflection of the ray $B'A'$ in $B'P$ meets $B'C'$ at E between B'' and C'. The

desired inequality now follows from Lemma 1 and

$$\begin{aligned} C'D + DA''B''E + EC' &\geqslant C'D + DA' + A'B' + B'E + EC' \\ &\geqslant C'A' + A'B' + B'C'. \end{aligned}$$

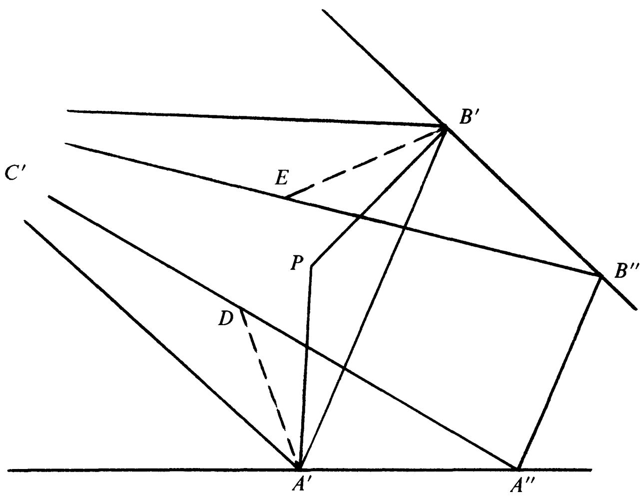


FIG. 6

THEOREM, Case 2b. See Figure 7. Suppose $a_2 \leqslant a_3$ and C' is on BA produced.

The proof is the same as 2a after one modification: the point E is defined as the point between B'' and C' where (it can be shown) $A'B'$ produced meets $B''C''$. □

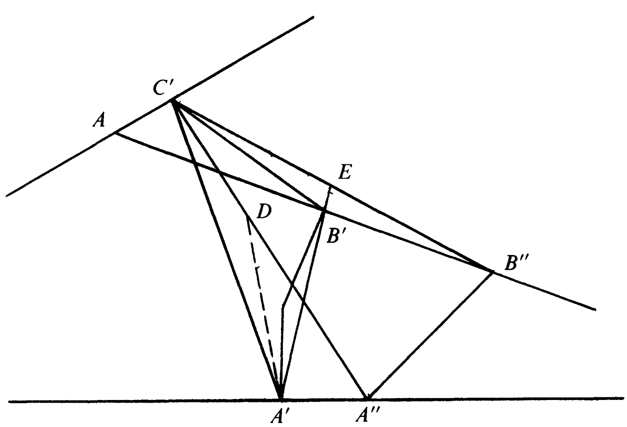


FIG. 7

This completes the proof of the proposed inequality. We note that it is clear from the proof that equality obtains if and only if the points A'', B'', C'' coincide with A', B', C' respectively—that is, if and only if the point P is the orthocenter.

A Recursive Sequence

E 2850. *Proposed by Clark Kimberling, University of Evansville.*

Let $a_1 = 1$, $b_1 = 2$, $c_1 = 3$. Beginning with $n = 1$ and continuing for $n \geqslant 2$, let $S(n)$ be the set of all a_i, b_i, c_i, for which $i \leqslant n$, and let

$$a_{n+1} = \text{least positive integer not in } S(n);$$
$$b_{n+1} = \text{least positive integer not in } S(n) \text{ and not } = a_{n+1};$$
$$c_{n+1} = a_{n+1} + b_{n+1}.$$

Let d_k be the increasing sequence of all n for which $b_n = a_n + 2$. Prove:

(i) $\lim_{k\to\infty} d_k/k = 6$;

(ii) If B is any integer, then $(d_k - 6k)/2 = B$ for infinitely many k.

Solution by D. M. Bloom, Brooklyn College of CUNY. We sketch a proof by induction on $n \geqslant 0$ that

$$\begin{array}{lll} a_{4n+1} = 10n+1; & b_{4n+1} = 10n+2; & c_{4n+1} = 20n+3 \\ a_{4n+2} = 10n+4; & b_{4n+2} = 10n+5; & c_{4n+2} = 20n+9 \\ a_{4n+3} = 10n+6; & b_{4n+3} = 10n+7; & c_{4n+3} = 20n+13 \\ a_{4n+4} = 10n+8+\delta; & b_{4n+4} = 10n+10; & c_{4n+4} = 20n+18+\delta, \end{array} \tag{1}$$

where $\delta = 0$ or 1 according to whether the exponent of 2 in $n+1$ is even or odd. Assuming that (1) holds up to and including a given n, we must show that (1) holds with n replaced by $n+1$. Only the values in the fourth row are not obvious. Let $T = \{c_1, c_2, \ldots, c_{4n+7}\}$; then

$$\begin{aligned} a_{4n+8} = 10n+19 &\Leftrightarrow 10n+18 \in T \Leftrightarrow 2 \text{ has even exponent in } (n/2)+1 \\ &\Leftrightarrow 2 \text{ has odd exponent in } n+2; \\ a_{4n+8} = 10n+18 &\Rightarrow 10n+18 \notin T \Rightarrow 10n+19 \in T \\ &\Rightarrow b_{4n+8} = 10n+20 \end{aligned}$$

from which the assertion regarding the fourth row of (1) is established.

It follows from (1) that the d's are precisely the integers of the form $4^k m$ where m is odd and $k \geqslant 1$. The number of such integers with fixed exponent k which do not exceed a fixed n is

$$s(k,n) = \left[(n+4^k)/2 \cdot 4^k\right] = n/2 \cdot 4^k + e(k,n)$$

(brackets denote the greatest integer function). Since $\Sigma_k(n/2 \cdot 4^k) = n/6$, the total number of d's not exceeding n is

$$s(n) = \sum_k s(k,n) = n/6 + E(n) = n/6 + \sum_k e(k,n).$$

Since $|e(k,n)| \leqslant \frac{1}{2}$ if $k \leqslant N = [\log_4(n)]$, and

$$\left|\sum_{k>N} e(k,n)\right| = n \sum_{k>N} (1/2 \cdot 4^k) < n/4^{N+1} < 1,$$

we have $|s(n) - n/6| < \frac{1}{2}(\log_4(n)) + 1 = o(n)$, so that $s(n)/n \to 1/6$ as $n \to \infty$. Hence $d_k/k = d_k/s(d_k) \to 6$ as $k \to \infty$, proving (i).

The integer $3 \cdot 4^n$ $(n \geqslant 1)$ is one of the d's, and $e(k, 3 \cdot 4^n)$ has the value 0 when $1 \leqslant k < n$, the value $\frac{1}{2}$ when $k = n$, and the value $-3/(2 \cdot 4^{k-n})$ when $k > n$. It follows that $E(3 \cdot 4^n) = 0$. Moreover, if $n > B \geqslant 0$, if h is either 0 or 1, and we let

$$M = 3 \cdot 4^n - (-1)^h(4^1 + 4^2 + \cdots + 4^B),$$

then for each k ($1 \leqslant k \leqslant B$) the number of integers $4^k m$ (m odd) between M and $3 \cdot 4^n$ (including M but not including $3 \cdot 4^n$) is

$$1 + (2 + 8 + 32 + \cdots + 4^{B-k}/2) = (4^{B+1-k} + 2)/6,$$

and M is one of the d's. Putting this result and that of the preceding sentence together, we see that $M = d_k$ where $(d_k - 6k)/2 = (-1)^h B$. Since for each B there are infinitely many choices of $n > B$, assertion (ii) follows.

A Gamma Function Inequality

E 2855. *Proposed by W. W. Meyer, General Motors Research Laboratories.*

For real nonnegative x, y, solve $\Gamma(x + y + 1) \geqslant (1 + xy)\Gamma(x + 1)\Gamma(y + 1)$.

Solution by the proposer. Define

$$f(x, y) = \frac{\Gamma(x + y + 1)}{\Gamma(x + 1)\Gamma(y + 1)}.$$

The gamma function has the following representation due to Euler [**1**], p. 237:

$$\Gamma(z + 1) = \prod_{n=1}^{\infty} [1 + 1/n]^z [1 + z/n]^{-1} \qquad (z \neq -1, -2, \ldots).$$

Convergence is absolute. Hence, the separate representations of $\Gamma(x + 1)$, $\Gamma(y + 1)$ and $\Gamma(x + y + 1)$ can be combined to yield

$$\begin{aligned} f(x, y) &= \prod_{n=1}^{\infty} \frac{[1 + 1/n]^{x+y}[1 + x/n]}{[1 + (x + y)/n]} \cdot \frac{[1 + x/n]}{[1 + 1/n]^x} \cdot \frac{[1 + y/n]}{[1 + 1/n]^y} \\ &= \prod_{n=1}^{\infty} \frac{[1 + x/n][1 + y/n]}{[1 + (x + y)/n]} \\ &= \prod_{n=1}^{\infty} \left[1 + \frac{xy}{n(n + x + y)}\right]. \end{aligned}$$

If we hold xy fast to a value $p > 0$, then f clearly becomes a monotone decreasing function of $s = x + y$. Since $(x, y) = (1, p) \Rightarrow s = 1 + p = f$, it follows that $f < 1 + p$ iff $s > 1 + p$. When $p = 0, f = 1$. So

$$f(x, y) < 1 + xy \Leftrightarrow 0 < xy < x + y - 1 \Leftrightarrow (1 - x)(1 - y) < 0 < xy.$$

That is to say, with the arguments x, y restricted to nonnegative reals, the inequality $\Gamma(x + y + 1) \geqslant (1 + xy)\Gamma(x + 1)\Gamma(y + 1)$ fails to hold true if and only if one of the arguments is greater than 1 while the other is less than 1 but positive.

Reference

1. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, Cambridge University Press, London, 1962.

Partially solved also by L. Kuipers (Switzerland).

Distance from a Hyperbola to an Ellipse

E 2857 [1980, 755]. *Proposed by John Leggett, Atascadero, California.*

If x, y, z, w are real numbers, $xy = 4$, $z^2 + 4w^2 = 4$, prove that $(x - z)^2 + (y - w)^2 \geqslant 1.6$.

Solution by U. Abel (Germany) and G. S. Lessells (Ireland) (independently). The tangent to $xy = 4$ at $(2\sqrt{2}, \sqrt{2})$, and the tangent to $z^2 + 4w^2 = 4$ at $(\sqrt{2}, \frac{1}{2}\sqrt{2})$ are parallel. They bound a strip of width $4/\sqrt{10}$.

Many solvers improved the bound 1.6 to 1.7747.... Some used Lagrange multipliers; some took an arbitrary point on the ellipse, and dropped a perpendicular to the hyperbola. The idea of drawing two parallel tangents is simple, but the calculations needed are not quite trivial.

Also solved by J. M. Ash, A. Bager (Denmark), D. M. Bloom, E. C. Buissant des Amorie (Netherlands), Chico Problem Group, C. Chouteau, J. E. Cruthrids & L. E. Mattics, R. Cuculière (France), G. A. Edgar, M. P. Eisner, N. Franceschine, W. J. Gorman, G. A. Heuer, P. Hohler (Switzerland), E. Johnston, F. Lanainger, D. Moore, M. Prendergast, I. A. Sakmar (Canada), I. J. Schoenberg, B. L. R. Shawyer (Canada), M. Bowe (Switzerland), M. Woltermann, and the proposer.

Unequal Trigonometric Sums

E 2874 [1981, 208]. *Proposed by Naoki Kimura and Tetsundo Sekiguchi, University of Arkansas.*

Let $n \geqslant 3$, $0 < A_i \leqslant \pi/2$, $i = 1, 2, \ldots, n$. Assume $\Sigma_{i=1}^{n} \cos^2 A_i = 1$. Prove

$$\sum \tan A_i \geqslant (n-1) \sum \cot A_i.$$

Solution by M. S. Klamkin, University of Alberta, and V. Pambuccian, student (Rumania) (independently). The inequality can be rewritten as

$$\sum \frac{1}{\sin A_i \cos A_i} \geqslant n \sum \frac{\cos A_i}{\sin A_i}. \tag{1}$$

Since by Cauchy's inequality,

$$\sum \sin A_i \cos A_i \sum \frac{1}{\sin A_i \cos A_i} \geqslant n^2,$$

(1) will follow from the stronger inequality

$$n \geqslant \sum \sin A_i \cos A_i \sum \frac{\cos A_i}{\sin A_i}. \tag{2}$$

By letting $x_i = \cos^2 A_i$, $S = \Sigma x_i$, (2) can be expressed in the homogeneous form

$$\frac{1}{n} \sum x_i \geqslant \left\{ \frac{1}{n} \sum \sqrt{x_i (S - x_i)} \right\} \left\{ \frac{1}{n} \sum \sqrt{x_i / (S - x_i)} \right\}. \tag{3}$$

Finally, we can assume $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$, so that $\{x_i(S - x_i)\}$ and $\{x_i/(S - x_i)\}$ are monotonic in the same sense and (3) follows by Chebysev's inequality with equality iff x_i are all equal. (See Hardy, Littlewood, and Pólya, *Inequalities*, pp. 43–44.)

REMARKS: By Cauchy's inequality again, (3) interpolates the power mean inequality

$$\frac{1}{n} \sum x_i \geqslant \left\{ \frac{1}{n} \sum \sqrt{x_i} \right\}^2.$$

Inequality (3) can be extended to

$$\frac{1}{n} \sum x_i \geqslant \left\{ \frac{1}{n} \sum x_i^r (S - x_i)^s \right\} \left\{ \frac{1}{n} \sum x_i^{1-r} (S - x_i)^{-s} \right\}$$

where $1 \geqslant r \geqslant s \geqslant 0$. Actually it even suffices to take $r/s \geqslant \max x_i/(S - x_i)$ (this ensures that $x^r(S - x)^s$ is an increasing function).

Also solved by B. Cheng & D. T. Hùng, L. Kuipers (Switzerland), J. Hrdina (Czechoslovakia), L. E. Mattics, St. Olaf Problem Solving Group, V. Peck and the proposers.

Kuipers showed that $n\sqrt{(n-1)}$ is at the same time a lower bound for the left member and an upper bound for the right member of the inequality in the proposal.

A Positive Real Outside the Intervals ($r_i - \varepsilon_n, r_i + \varepsilon_n$) for the Positive Rationals

E 2881 [1981, 291]. *Proposed by Nick Franceschine III, Sebastopol, Calif.*

Enumerate the positive rationals as Cantor did, $r_1 = 1, r_2 = 1/2, r_3 = 2, r_4 = 3, r_5 = 1/3, r_6 = 1/4, r_7 = 2/3, \ldots$. Exhibit an infinite set of real numbers outside every one of the closed intervals $[r_n - 2^{-n-1}, r_n + 2^{-n-1}]$.

Solution by University of South Alabama Problem Group. The key word is "exhibit," since clearly most positive reals lie outside all the intervals.

Let $\rho = (\sqrt{5} + 1)/2$ and c be any nonnegative integer, we will show that $c + \rho$ is not in any of the given intervals.

For any n, set $r_n = a_n/b_n$ with $(a_n, b_n) = 1$. If $a_n + b_n = m + 1$ for $m \geqslant 2$, then for any i with $2 \leqslant i \leqslant m - 1$ there are at least two r_j with $a_j + b_j = i + 1$, namely i and $1/i$ so $n > 2(m - 2) + 1$. Hence, for $m \geqslant 4$,

$$4b_n^2 \leqslant 4m^2 \leqslant 2^{2(m-2)+3} \leqslant 2^{n+1}$$

and we can now check to show that if n is not 2 then $4b_n^2 \leqslant 2^{n+1}$.

If for some $n, |r_n - (c + \rho)| < 2^{-(n+1)}$, then n is not 2 so r_n is a convergent of $c + \rho$. So b_n is F_i the ith Fibonacci number for some i. But then we have from the theory of continued fractions that $(F_i F_{i+2})^{-1} < |r_n - (c + \rho)|$ which would imply $F_i F_{i+2} > 4F_i^2$ which does not hold. We are done.

The proposer and Russell Lyons, student, University of Michigan, exhibited the set $k + 1/\sqrt{m}$, for all nonsquare integers m and all positive integers k and pointed out that "positive" was inadvertently omitted from the proposal.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by February 28, 1983. The solver's full post-office address should be on each sheet.

6399. *Proposed by Armel Mercier, Université du Québec à Chicoutimi.*

For all integers $N \geqslant 2$ prove that

$$\sum_{n=2}^{N} \left(\frac{(-1)^{n+1}\binom{N}{n}}{n} \sum_{k=1}^{n-1} (-1)^k \binom{n}{k+1} \zeta(1+k) \right) = \log N + o(1),$$

where $\zeta(s)$ stands for the Riemann zeta function.

6400. *Proposed by Fred Galvin, University of Kansas, and Karel Prikry, and J. Ian Richards, University of Minnesota.*

Let $2 \leqslant k \leqslant n$. A $k \times n$ Latin rectangle is a $k \times n$ matrix with entries $1, 2, \ldots, n$, such that no row or column contains a repetition. Let $R(k, n) = R^+(k, n) - R^-(k, n)$ where $R^+(k, n)$ $[R^-(k, n)]$ is the number $k \times n$ Latin rectangles whose first two rows are of the form

$$\begin{matrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{matrix}$$

where π is an even [odd] permutation of $\{1, 2, \ldots, n\}$.

(i) Show that $R(2, n) = (-1)^{n-1}(n - 1)$.
(ii) Show that $(-1)^{n-1}R(3, n) \geqslant 0$; the inequality is strict for $n > 4$.
(iii)*Is $(-1)^{n-1}R(k, n) \geqslant 0$ for $k \geqslant 4$? In particular, is $(-1)^{n-1}R(n, n) \geqslant 0$ for all n?

6401. *Proposed by J. C. Lagarias, Bell Laboratories, Murray Hill, NJ, and D. S. Sturtevant, Massachusetts Institute of Technology.*

Let $A = \{a_1, \ldots, a_n\}$ where $1 \leqslant a_1 < a_2 < \cdots < a_n$ are positive integers. Let r be a given positive rational number, with $r = p/q$ and $(p, q) = 1$. Define

$$d_r(A) = \#\left\{(i, j, k) \colon i < j < k \text{ and } \frac{a_j - a_i}{a_k - a_j} = r\right\}.$$

Now let A vary and set

$$d_r(n) = \max_A \{d_r(A)\}.$$

Prove or disprove that $d_r(n) \sim \dfrac{n^2}{2(p+q)}$ as $n \to \infty$.

6402. *Proposed by the Editors.*

Prove that if $n \geqslant 8$, the alternating group $\mathcal{Q}_n$ can be generated by two conjugate four-groups.

SOLUTIONS OF ADVANCED PROBLEMS

Derivatives of Continuous Functions

6140* [1977, 221; 1981, 296]. *Proposed by F. S. Cater, Portland State University, Oregon.*

Let f be a continuous real-valued function on [0, 1] and let E_f denote the (possibly void) set $\{x \in [0, 1] \colon f'(x)$ exists and is finite$\}$. Let $a(f) =$ Lebesgue outer measure of $f([0,1] \setminus E_f)$, and let

$$m(t) = \begin{cases} f'(t) & \text{for } t \in E_f \\ 0 & \text{otherwise.} \end{cases}$$

Let $b(f) = a(f) + \int_0^1 m(t)\,dt$ and $c(f) = a(f) + \int_0^1 [1 + m(t)^2]^{1/2}\,dt$.

Find (1) $\max c(f)$ over all f such that $b(f) = 1$, and (2) $\min c(f)$ over all f such that $b(f) = 1$. Describe functions for which $c(f)$ takes one of these values.

Further comments by K. F. Andersen, University of Alberta, F. S. Cater, Portland State University, and M. J. Pelling, London, England. By Minkowski's inequality for integrals [see, for example, Hardy, Littlewood, and Pólya, Inequalities, p. 146],

$$c(f) = a(f) + \int_0^1 \left[1 + m(t)^2\right]^{1/2} dt$$

$$\geqslant a(f) + \left[\left(\int_0^1 1\,dt\right)^2 + \left(\int_0^1 m(t)\,dt\right)^2\right]^{1/2} \tag{1}$$

with equality if and only if $m(t) = k$ a.e., k constant. Now $b(f) = 1$ yields

$$c(f) \geqslant a(f) + \left\{1 + [1 - a(f)]^2\right\}^{1/2} \tag{2}$$

and for $a(f) \geqslant 0$, the term on the right-hand side of (2) is increasing in $a(f)$; hence

$$c(f) \geqslant \sqrt{2}, \tag{3}$$

with equality in (3) if and only if $a(f) = 0$ and $m(t) = k$ a.e.; moreover, $k = 1$ in this case since $b(f) = 1$. But if $a(f) = 0$ and $m(t) = 1$ a.e., then for any interval J the Lebesgue outer measure λ satisfies $\lambda(f(E_f \cap J)) \leqslant \lambda(E_f \cap J) \leqslant \lambda(J)$ [see, for example, E. Hewitt & K. Stromberg, *Real and Abstract Analysis*, Exercise 17.27, p. 269] and (from $a(f) = 0$) $\lambda(f(J)) \leqslant \lambda(J)$; from this we also see that f is absolutely continuous and $f(t) = t + c$ for some constant c. Such functions f and only such functions minimize $c(f)$. On the other hand, the requirement $b(f) = 1$ does not

guarantee that $c(f)$ is bounded as the example $f_n(t) = \sin(2n + \frac{1}{2})\pi t$ clearly shows, since $b(f_n) = 1$ while $c(f_n)$, the arc length of the graph of $f_n(t)$, certainly exceeds $2n$.

Pelling refers to D. E. Barburg: "On Differentiable Transformations in R^n," Papers in Analysis, Supplementary Volume to the MONTHLY for April 1966.

n Points in the Plane, ϵn Distinct Distances

6323 [1980, 826]. *Proposed by P. Erdös, Hungarian Academy of Sciences.*

Let k, n be integers.

(i) Let $n - 1 \leqslant k \leqslant \frac{1}{2}n(n-1)$. Prove that there exist n distinct points $x_1, x_2, \dots, x_n$ on the line that determine exactly k distinct distances $|x_i - x_j|$.

(ii) Let $[\frac{1}{2}n] \leqslant k \leqslant \frac{1}{2}n(n-1)$. Prove there are n distinct points in the plane that determine exactly k distinct distances.

(iii) *Is it true that, for every $\epsilon > 0$, there is an $n_0 = n_0(\epsilon)$ such that, if $n > n_0$ and if $\epsilon n < k \leqslant \frac{1}{2}n(n-1)$, there are n points in the plane that determine exactly k distances?

Solution by Don Coppersmith, IBM Research, Yorktown, Heights, NY.

(i) Let m be an integer between 1 and $n - 1$ such that

$$n(n-1)/2 - (m-1)(m-2)/2 \geqslant k \geqslant n(n-1)/2 - m(m-1)/2.$$

Let the first m of our points be $1, 2, \dots, m$. Let the next point be $m + p$, where $p = k - (m-1) - n(n-1)/2 + (m+1)m/2$, so that $1 \leqslant p \leqslant m$. The distances between these first $m + 1$ points are precisely the set $\{1, 2, \dots, p + m - 1\}$. Now place the other $n - m - 1$ points in "generic" position, so that no distances are duplicated; e.g., the new points can be $x_i = \pi^i$, $m + 2 \leqslant i \leqslant n$. Thus the total number of distinct distances is

$$(p + m - 1) + \left(\sum_{i=m+2}^{i=n} (i-1) \right) = k.$$

(ii) If $[n/2] \leqslant k < n - 1$, let the n points occupy consecutive vertices of a regular $2k + 1$-gon. There are exactly k different distances among the vertices of this $2k + 1$-gon, and all are exhibited. If $k \geqslant n - 1$, imbed the line in the plane and apply part (i).

(iii) We give a construction, for large enough n, showing each number k of distinct distances, for $\varepsilon n \leqslant k \leqslant n/2$. Combine with (ii) to complete the solution.

Let $m = 3 + [\sqrt{n}]$. For $m \leqslant a \leqslant n - 2m$, $1 \leqslant b \leqslant m$, let our set be the following subset of the integer lattice:

$$\{(i,1) \mid 1 \leqslant i \leqslant a\} \cup \{(1,j) \mid 2 \leqslant j \leqslant m\} \cup \{(0,h) \mid 1 \leqslant h \leqslant b\} \cup C$$

where C is any subset of $\{(i,j) \mid 2 \leqslant i \leqslant a - 1, 2 \leqslant j \leqslant m\}$ of the proper size. (We have chosen a, b, and m such that this is possible.) Now observe that the squared distances between points are exactly the set

$$\{i^2 + j^2 \mid 0 \leqslant i \leqslant a - 1, 0 \leqslant j \leqslant m - 1\} \cup \{a^2 + j^2 \mid 0 \leqslant j \leqslant b - 1\} - \{0\}.$$

Let $D(a, b)$ be the set of *distinct* squared distances so obtained. Then notice that

$$D(m,1) \subset D(m,2) \subset \cdots \subset D(m,m) \subset D(m+1,1) \subset D(m+1,2)$$
$$\subset \cdots \subset D(m+1,m) \subset \cdots D(n-2m,1) \subset \cdots \subset D(n-2m,m)$$

is a sequence of nondecreasing sets: each is obtained from the previous by adding one squared distance, which may or may not duplicate a squared distance already in the previous set. So the cardinalities $|D(a,b)|$ form a nondecreasing sequence, each term of which differs from the previous by 0 or 1. Thus the cardinalities exhaust all the integers between the lower and upper bounds $|(D(m,1)|$ and $|D(n-2m,m)|$, respectively).

The upper bound $|D(n-2m, m)|$ is at least $n - 2m - 1 \geqslant [n/2]$. The set $D(m, 1)$, leading to the lower bound, is a set consisting only of integers, between 1 and $2m^2 \sim 2n$, each of which is the sum of two squares. It is well known that the sum of two squares cannot be divisible by any prime $p \equiv 3 \bmod 4$ unless it is divisible by p^2. Thus the asymptotic density of integers expressible as the sum of two squares is bounded by $\Pi(1 - 1/p + 1/p^2)$, the product being taken over all primes $p \equiv 3 \bmod 4$. The indicated product tends to 0. Thus, for n large enough, $|D(m, 1)| < \epsilon(2n)$, as required.

Also solved by L. Kuipers (Switzerland), C. Löfgren (Sweden), O. P. Lossers (The Netherlands), H. M. Marston, J. Ward, and the proposer.

Product of Quasi-connected Sets

6326 [1981, 68]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

A topological space S is *quasi-connected* provided that any covering of S by disjoint nonempty open sets is finite. Equivalently, any continuous map from S into a discrete space has finite range. It is not hard to show that, if S is quasi-connected and T is either connected or compact, then $S \times T$ is quasi-connected.

Is the product of two quasi-connected spaces always quasi-connected? What about infinitely many?

Solution by Dennis K. Burke, Miami University. The product of two quasi-connected spaces is not always quasi-connected. First observe that any countably compact space is quasi-connected. It is well known that there exist countably compact subspaces X and Y of βN (the Stone-Cech compactification of the natural numbers) such that $X \cap Y = N$ (J. Dugundji, *Topology*). For any such pair, the set $\Delta = \{(n, n): n \in N\}$ is closed in $X \times Y$ and each singleton $\{(n, n)\}$, for $n \in N$, is open in $X \times Y$ so the collection $U = \{X \times Y - \Delta\} \cup \{\{(n, n)\}: n \in N\}$ is a covering of $X \times Y$ by an infinite number of disjoint open sets.

Andreas Blass, University of Michigan, gave essentially the same argument in a self-contained form. The following solvers gave references to examples showing that the product of two quasi-connected spaces is not always quasi-connected: K. P. Hart (Netherlands), John Isbell, David Jakel, A. Le Donne (Italy) and R. Wilson (Mexico), Peter J. Nyikos, and Erik van Douwen. Their references are as follows:

1. H. Terasaka, On Cartesian product of compact spaces, Osaka J. Math. 4 (1952) 11–15.

2. J. Novak, On the Cartesian product of two compact spaces, Fund. Math., 40 (1953) 106–112.

3. Gilman and Jerison, Rings of Continuous Functions, Van Nostrand, Princeton, NJ, 1960, p. 135 (example 9.15).

4. R. S. Pierce, Rings of continuous integer-valued functions, Trans. Amer. Math. Soc., 100 (1961) 371–394.

Also solved by Ronnie Levy.

The Equation per $(AX) = \text{per}(A)\text{per}(X)$

6330 [1981, 149]. *Proposed by Ko-Wei Lih, Academia Sinica, Taiwan.*

Characterize all $n \times n$ complex matrices A for which the relation per $(AX) = \text{per}(A)\text{per}(X)$ holds for every complex $n \times n$ matrix X. (Here $\text{per}(B)$ denotes the permanent of B.) Does the same characterization hold when the complex field is replaced by an arbitrary field?

Solution by Miroslav D. Ašić, Department of Mathematics, The London School of Economics. Let us consider the following two types of square matrices of order n:

(i) matrices with at least one zero-row;

(ii) matrices which have exactly one nonzero entry in each row and each column. It is easily

seen that these two types of matrices do satisfy the condition

$$(*) \qquad \mathrm{per}(AX) = \mathrm{per}(A)\mathrm{per}(X), \quad \text{for every matrix } X \text{ of order } n.$$

We shall show that those are the only matrices which satisfy (*). Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

and take

$$X = \begin{bmatrix} t & \dots & t \\ \vdots & & \vdots \\ t^n & \dots & t^n \end{bmatrix}.$$

The condition (*) is in this case

$$n!\,\mathrm{per}(A)t^{n(n+1)/2} = n!p_1(t)\cdots p_n(t)$$

where

$$p_i(t) = a_{i1}t + \cdots + a_{in}t^n, \qquad i = 1,\dots,n.$$

If $\mathrm{per}(A) = 0$, then at least one of the polynomials p_i is identically zero, i.e., A contains at least one zero-row. If $\mathrm{per}(A) \neq 0$, then each polynomial p_i has exactly one nonzero term, which means that the matrix A is of the type (ii).

If the complex field is replaced by an arbitrary field, the answer may be different. For instance, if we take a field of characteristic two, then obviously $\mathrm{per}(B) = \det(B)$ for each square matrix B so that (*) holds for each square matrix A.

Also solved by LeRoy B. Beasley, Benny Cheng and Dinh Thế Hùng, Marvin Marcus, Frank W. Schmidt & Rodica Simion, and the proposer.

Comment. Marcus gave a complete analysis of the problem in the context of the symmetric algebra of an n-dimensional vector space V over a field F. If $V^{(m)}$ and $V^{*(m)}$ are respectively the mth completely symmetric spaces over V and its dual V^*, then these spaces are dual with respect to the pairing given by the permanent function. Under the condition that $\mathrm{char}(F) > m!$, $m \leqslant n$, he obtains the set of A on which the permanent function is multiplicative for all X. He noted that his proof carries over virtually unaltered when the permanent function is replaced by

$$d_H(Y) = \sum_{\sigma \in H} \prod_{i=1}^{m} Y_{i\sigma(i)},$$

for H a subgroup of the symmetric group S_m, and the symmetric space is replaced by the symmetry class of tensors defined by H. Marcus further noted that some work on maximal subgroups of the general linear group on which the permanent is multiplicative is discussed in H. Minc, Permanents, Encyclopedia of Mathematics and its Applications, volume 6, Addison-Wesley, 1978, p. 153.

Primary Decomposition of Ideals

6335 [1981, 213]. *Proposed by Harley Flanders, Florida Atlantic University.*

In the ring $R = C[X, Y, Z]$, find the primary decomposition of the ideal

$$A = (X^3 - Y^2, Y^3 - Z^2, Z^3 - X^2).$$

Solution by Barbara L. Osofsky, Rutgers University. Let

$$R=\mathbb{C}[x,y,z],\, I=\langle x^3-y^2,y^3-z^2,z^3-x^2\rangle\subseteq R,$$

where $\langle S\rangle$ denotes the ideal generated by the set S. Let $\alpha=e^{2\pi i/19}$ or any other nonreal 19th root of 1. Then the primary decomposition for $I\subseteq R$ is

$$I=\langle x^2,y^2,z^2\rangle\cap\bigcap_{k=0}^{18}\langle x-\alpha^k,y-\alpha^{11k},z-\alpha^{7k}\rangle.$$

This is proved by the observations:

(i) Each of the indicated ideals is primary. Indeed, all but $\langle x^2,y^2,z^2\rangle$ are maximal, and $\langle x^2,y^2,z^2\rangle$ is primary for $\langle x,y,z\rangle$ since that is the only prime containing it.

(ii) Each of the indicated ideals contains I since $\alpha^{3k}=\alpha^{22k}$, $\alpha^{33k}=\alpha^{14k}$, and $\alpha^{21k}=\alpha^{2k}$.

(iii) A $\mathbb{C}$-basis for $R/\langle x^2,y^2,z^2\rangle$ is

$$\{1,x,y,z,xy,xz,yz,xyz\}$$

so by the Chinese Remainder Theorem, R modulo the intersection in the right hand side of the equation has C-dimension $8+19=27$.

(iv) A $\mathbb{C}$-basis for R/I is $\{x^iy^jz^k\,|\,0\leqslant i,j,k\leqslant 2\}$ since any monomial containing the cube of x,y, or z is congruent modulo I to a monomial of smaller degree. Hence the $\mathbb{C}$-dimension of R/I is 27.

The decomposition was arrived at by noting that (iv) implies that R/I is artinian so every prime containing I is maximal. Since $\mathbb{C}$ is algebraically closed, maximal ideals are of the form $M=\langle x-a,y-b,z-c\rangle$ where $a,b,c\in\mathbb{C}$. If $M\supseteq I$, ignoring non-well-definedness,

$$b=a^{3/2},\qquad c=b^{3/2}=a^{9/4},\qquad a=c^{3/2}=a^{27/8},$$

so $a=0$ or $a^{19/8}=1$ and $a^{19}=1$. Moreover, the triples (a,b,c) which occur in maximals containing I are closed under coordinate-wise multiplication. This leads to the relation

$$\bigcap_{k=0}^{18}\langle x-\alpha^k,y-\alpha^{11k},z-\alpha^{7k}\rangle\supseteq I.$$

The component $\langle x^2,y^2,z^2\rangle$ comes from the observation that

$$(x^2-z^3)+(z^2-y^3)z+(y^2-x^3)yz=(1-xyz)x^2\in I$$

and by symmetry, $(1-xyz)\langle x^2,y^2,z^2\rangle\subseteq I$ making the primary ideal $\langle x^2,y^2,z^2\rangle$ look like the annihilator of an element in R/I. The counting in (iii) and (iv) simply confirmed that.

A change of fields in the above can be made. Let $R'=\mathbb{Q}[x,y,z]$ and

$$I'=\langle x^3-y^2,y^3-z^2,z^3-x^2\rangle\subseteq R'.$$

Then the primary decomposition for I' is

$$I'=\langle x^2,y^2,z^2\rangle\cap\langle x-1,y-1,z-1\rangle\cap\left\langle\sum_{i=0}^{18}x^i,1-xyz,I'\right\rangle.$$

The first two components have generators as in R, and the last is generated by a set of generators for

$$\bigcap_{i=0}^{18}\langle x-\alpha^k,y-\alpha^{11k},z-\alpha^{7k}\rangle\subseteq R.$$

The lack of symmetry can be eliminated by a computation which shows that, modulo $\langle 1-xyz,I'\rangle$,

$$\sum_{i=0}^{18}x^i=\sum_{\substack{0\leqslant i,j,k\leqslant 2\\ ijk=0}}x^iy^jz^k.$$

Moreover,

$$R'\Big/\left\langle \sum_{i=0}^{18} x^i, 1 - xyz, I' \right\rangle$$

has $\mathbb{Q}$-dimension 18 and contains a copy of $\mathbb{Q}[\alpha]$ so $\langle \Sigma_{i=0}^{18} x^i, 1 - xyz, I' \rangle$ is maximal in R'.

Also solved by Marshall Fraser, Robert Gilmer & Watten Nichols, and the proposer. Gilmer and Nichols determine the primary decomposition in $\mathbb{C}[X_1, \ldots, X_n]$ of $(X_1^a - X_2^b, X_2^a - X_3^b, \ldots, X_n^a - X_1^b)$ where $(a, b) = 1$.

Embedding One Metric Space in Another

6343 [1981, 295]. *Proposed by W. Holsztynski, Ann Arbor, Michigan.*

(a) Show that there exists a four-element metric space which cannot be isometrically embedded in a Hilbert space.

(b) Let X be the Banach space obtained from $\mathbb{R}^2$ by defining the norm $\|(x, y)\| = \max\{|x|, |y|\}$. Show that every four-element metric space can be isometrically embedded in X.

Solution by R. C. James and R. A. Vitale, Claremont Graduate School, Claremont, California. (a) Let the four-element metric space be $\{P_1, P_2, P_3, P_4\}$ and denote $d(P_i, P_j)$ by d_{ij}. Suppose $d_{24} = 2$, d_{13} is any number for which $0 < d_{13} \leqslant 2$, and $d_{ij} = 1$ otherwise $(i \neq j)$. Since $d_{24} = d_{21} + d_{14}(2 = 1 + 1)$, an isometric embedding into a Hilbert space must send P_1 to the midpoint of the line segment determined by the images of P_2 and P_4. By similar reasoning, P_3 is sent to the same point, violating $d_{13} > 0$.

The same argument applies to any *strictly convex* normed linear space, i.e., a normed linear space for which $\|u + v\| = \|u\| + \|v\|$ only if one of u or v is a multiple of the other.

(b) Let the points of an arbitrary four-element metric space be labeled so that $d_{12} + d_{34} = \max\{d_{12} + d_{34}, d_{23} + d_{14}, d_{24} + d_{13}\}$, and then let

$$\begin{aligned} &P_1 \leftrightarrow (0,0), \\ &P_2 \leftrightarrow (d_{12}, d_{23} + d_{14} - d_{34}), \\ &P_3 \leftrightarrow (d_{13}, d_{14} - d_{34}), \\ &P_4 \leftrightarrow (d_{12} - d_{24}, d_{14}). \end{aligned}$$

Then $d_{12} = \|(d_{12}, d_{23} + d_{14} - d_{34})\|$, since $d_{23} + d_{14} - d_{34} \leqslant d_{12}$ by assumption, and $-d_{23} - d_{14} + d_{34} \leqslant d_{12}$ follows from $d_{34} \leqslant d_{41} + d_{12} + d_{23}$. Similarly, $d_{34} = \|(d_{12} - d_{24} - d_{13}, d_{34})\|$. Also, $|d_{12} - d_{13}| \leqslant d_{23}$ implies $d_{23} = \|(d_{12} - d_{13}, d_{23})\|$, and the remaining equalities follow as easily.

Also solved by Miroslav D. Asic (England), Lilian G. Button and D. Hammond Smith (England), Dinh Thế Húng and Benny Cheng, Dorothy Wolfe, and the proposer.

---

84. MISCELLANEA

LINEAR EQUATIONS

Determinant none,
Solution: lots or none.
Determinant some,
Solution: just one.

—Franklin Kemp

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

A Brief Course of Higher Mathematics. By V. A. Kudryavtsev and B. P. Demidovich. Moscow, Mir, 1981. 693 pp.

PETER J. HILTON
Mathematical Research Institute, ETH, Zürich, Switzerland

AMERICAN EDUCATION VERSUS SOVIET EDUCATION: THE BALANCE OF ERROR

Intensive efforts have been made, notably by Professor Izaak Wirszup of the University of Chicago, the best-informed person in this country on Soviet scientific education, to alert the government and people of the United States to the gross disparity in the level of preparation in mathematics and related sciences of people in comparable situations in the countries described, with no intended irony, as the two superpowers. The response of the Reagan administration cannot be regarded as too encouraging. Funding for the Science Education Directorate of the National Science Foundation has been cut to the bone, and beyond; and it is now reliably reported that research grants in pure science will be reduced, even in absolute dollar amounts. More generally, education in the United States is everywhere feeling the pinch. Class sizes are increasing; more part-time teachers are being employed; fewer young people are entering the teaching profession, being discouraged by uncertain career prospects and, in the case of the scientifically trained, encouraged by advantageous salary differentials to look elsewhere for employment. "Supply-side economics," the highly misleading but very modish catch-phrase of the faceless gurus of the Washington establishment, may be translated, for those concerned with education, into a system and a philosophy favoring short-term advantage and highly prejudicial to the most long-term of all social policies, educational progress.

But even if there were not such glaring disparities, at the political and social level, between attitudes to and support for scientific education in the United States and the Soviet Union, there would remain differences of educational philosophy which would be having their effect. It is with these differences that the rest of this article will be concerned, but to have failed even to mention those other disparities would have rendered the author liable to the charge, so damaging today, of academicism.

A distinguishing feature, then, of American and Russian education is the insistence, in this country, on breadth of curriculum right up through a four-year (or two-year) undergraduate program. In this respect, the Russian system is firmly embedded in the European tradition, so that, in order to understand that system, it is necessary to deal extensively with this tradition, which is highly selective, identifying at the (relatively) tender age of 12 or 13 the child likely to benefit from an academic education, and resolving, at the age of 16 or 17, the principal direction of a student's studies. Thus a student already earmarked for a university education will characteristically decide, in the early teens, to study mathematics and related disciplines and will spend the last 2 or 3 years in the secondary school pursuing a fairly narrow course of study centered on mathematics. He, or she, will enter university far better prepared in mathematics than his opposite number in America; and the discrepancy will increase throughout the undergraduate years. Here, however, a distinction must be made between those European countries in which the universities offer a bachelor's degree and those in which the students seek professional qualifications, or degrees at the master's or doctoral level. Moreover even where, as in the United Kingdom, a bachelor's degree is offered, the degree course tends to be shorter than in the United States—in the case of the United Kingdom, it is a 3-year course.

It would require an analysis far more profound and more thorough than could be accommo-

dated in these pages to reach a substantive conclusion on the relative merits of the two systems. First of all, criteria would have to be established for such a judgment; and, such is the role of education in our lives, the elaboration of such criteria would involve a discussion on the nature and purpose of human life itself. We must therefore content ourselves with a discussion of some of the effects of the two systems.

As a first approximation, we may say that the American system aims at breadth, and the European system at depth. Moreover, if we assume that American and European students have comparable capacities and spend roughly the same amount of time in study, then the American system achieves breadth at the expense of depth and the European system achieves depth at the expense of breadth. This means that American students, in any given academic year, learn less mathematics than their European counterparts, but devote time, not available to the European student, to outside subjects. If this time were well spent, it would be possible to argue that there would be a positive effect on the learning of mathematics—even if the student were not acquiring mathematical skills and techniques at maximum rate—since the role of mathematics would be better understood by a student of broader interests and attainments. However, experience leads one to doubt whether this is, in fact, what typically happens. The "breadth" which the American student acquires tends often to be nothing more than the result of exposure, at a rather superficial level, to a number of unrelated disciplines, so that the interactive aspect of a broad curriculum, which must be its principal justification, is all too often absent.

On the other hand, European—and, more especially, Soviet—students may also benefit less than one would expect from the greater rate of progress available to them through their more concentrated curriculum. For a concomitant of this concentration is early specialization, and for many students that specialization will be manifestly premature. Students may decide their area of specialization on the basis of the relative prestige of various disciplines prevailing in society (the application of this rule to Soviet society is obvious, but it applies also, for example, to French society); they may decide on the basis of their, or their parents', perception of the job market; they may decide on the basis of the popularity of a particular teacher; or they may decide on the basis of the subjects available to them in their particular secondary school. This last possibility has a conspicuous effect in the United Kingdom, where students are admitted to particular university courses rather than to the university as a whole; and where there is a tendency for students to apply to departments in the sciences and the arts because those departments offer them what they enjoy studying, and are good at, whereas they apply to engineering and social science departments often because they want to get away from what they were doing at secondary school.

Still talking very approximately, we may say that the American system favors the average student, and the late developer, while the European system favors the exceptionally bright student, especially if that brightness reveals itself early and assumes a very definite direction. This distinction seems to accord with certain aspects of the social ideals traditionally associated with the United States and Western Europe. For the United States is—or, rather, was intended to be—an egalitarian society, whereas Western European democracy has always accommodated itself to the class divisions among the inhabitants of the countries in which it has flourished. Thus it is not so surprising that in Western Europe one seeks to maximize the successes, while in the United States one seeks to minimize the failures. That the Soviet Union follows the Western European model in this respect should cause no surprise to those familiar with the extraordinary perquisites of the successful and the powerful in that least egalitarian of all contemporary societies.

Plainly there are many highly significant consequences that flow from these two diverging philosophies of education. The European model implies a highly selective intake into the university and strong competition for admission to graduate schools. The American model implies that post-secondary education be available, effectively, to all. The difference in cost alone is enormous. But there are other aspects of the American model to which attention should be drawn.

Egalitarianism implies equality of opportunity and, in particular, of educational opportunity.

But education itself, if successful, does not produce equality—it enhances differences between individuals, bringing out their wonderful uniqueness. American education, in practice, seems sometimes to be concerned, however perversely, to homogenize those who enjoy its benefits, and to count it a success if individual differences and individual attributes become blurred in the educational process. In particular, the education of the talented seems to be regarded by many as a matter of low priority. Let us agree that it must not detract from the admirable and essential social process of helping, through education, disadvantaged minorities, and women, to overcome the handicaps society has unfairly imposed upon them; but it cannot be that a great and powerful nation lacks the resources to do both. Only those who fail totally to understand the long-term interests of our society—unfortunately, this seems to include many people influential in decision-making in this country—could countenance the neglect of the talented.

If failures are to be minimized, frequent monitoring is required. This leads, in the U.S., to the prevalence of testing as a ubiquitous feature of all education, far more so than in the European model. With testing come the twin evils of teaching for tests and learning for tests, about which the readers of this journal will already be fully alerted and thoroughly informed. Moreover, the same philosophy leads to the manifestation of great care and concern that the students do as well as possible; but this fine aspiration leads in its turn to the failure to treat students as adults, and to the devising of means of preventing the less scrupulous from gaining unfair advantage over their more honest cohorts. The result, so far as mathematics is concerned, is to conduct ridiculously frequent tests, under conditions which contradict all we know of the optimal circumstances under which mathematics is done. Studying within a system dominated by tests, the student comes to believe that mathematics is something you *learn*, and that you only *do* it in order to convince the authorities that you have indeed learnt it. Success as a student becomes synonymous with success in tests; and the students devote much effort to trying to obtain from their teachers guarantees that the tests will be "fair"—that is, stereotyped and requiring only memory, not thinking. In this respect they share a common interest with the faculty who cannot afford a high failure rate. It is inevitable that, if the failures are to be minimized, there will be a tendency, unconscious in some cases, to simplify the syllabus and to adjust the pass level.

This lugubrious litany of defects in the practice of education in the United States is quite unfairly one-sided. The elitism present in the European system surely has its own unfortunate consequences, too. We have already spoken of premature specialization; we could have added unavoidable specialization, and the consequence that a student failing in a chosen specialty must leave the university, not having the option of several different majors. More complementary to the defects produced by too frequent and imperfectly structured tests are the consequences of the laissez-faire approach to education at, say, a West German university where a student follows his, or her, own program and may well have no contact with a professor whatever. It is the students' responsibility to decide when they are ready for their examination; and it is far from unknown for a student to linger on at a university for some years and then to quietly disappear, preferring oblivion to the humiliation of an unsuccessful examination.

One could continue in this vein to deal with generalities of educational philosophy within the two systems, but it were better to devote the rest of this essay to more specific questions of the mathematical content of the curriculum. As we have said, that content is richer in the European system than in ours. The balance, however, is redressed at the graduate level, where the American program offers a wonderful richness of content and opportunity for the student to gain maturity as an apprentice mathematician. Typically, a British graduate student starts working for the Ph.D. immediately following the completion of the bachelor's program (indeed, the term "graduate student" is not used; such a person is a "research student"). In the Soviet Union the doctorate is a very advanced degree indeed, to be gained only by mature and established scholars, so the correct comparison is with the "candidate" degree. Thus there is no suggestion that the professional mathematician receives an inferior training in the United States compared with his European counterpart—some would indeed argue the contrary proposition that the American mathemati-

cian learns more mathematics. Can we say anything, in general terms, about *what* mathematics is emphasized in the two systems?

Generalizations, in this field, are dangerous, but texts such as Kudryavtsev and Demidovich do tell us something about the typical content of a Soviet curriculum. Here we have *A Brief Course of Higher Mathematics* (the translation is accurate) which must cause the mind to boggle at what a full-size course would be! The intended student may be in any of the "natural, applied and engineering sciences." The coverage goes way beyond that of the first two years of an American university curriculum—the core program—and the treatment is rigorous. There is a long chapter on probability and statistics and a short chapter on linear programming; otherwise there is very little "applied mathematics" as we understand the term. Linear algebra is confined to 3 dimensions until the final chapter (on linear programming). There is no abstract algebra and no notion of topology, but much on infinite series and differential equations.

It appears that the structure of courses in the Soviet system, as in the British system, militates against taking a broad view of what is good mathematics and, more especially, what is good applied mathematics. Applied mathematics is, in those systems, almost synonymous with theoretical physics; and mathematics which has not yet been applied to physics and engineering finds no place in the curriculum, which is therefore very traditional. In particular, there is no indication that combinatorics deserves any place in the general education of future users of mathematics, nor that there needs to be any discussion of the general principles governing the effective use of mathematics—how, for instance, one chooses the appropriate mathematical model.

It is striking that, whereas Soviet institutions are—officially— based on a materialist philosophy, the curriculum here elaborated would, in American terms, appear very "pure"—even for a mathematics major. Perhaps Soviet mathematicians understand that, in order to be able to apply mathematics effectively, it is necessary to have a firm grasp of a great deal of mathematics; and perhaps a less brief course (this text has a mere 693 pages) would include some more modern mathematics and more applications, taken from a broader spectrum of sciences. Perhaps, too, Soviet mathematicians can sufficiently establish their political and philosophical orthodoxy by prefacing their texts with the charmingly irrelevant references to Friedrich Engels and to the contributions of Russian mathematicians which grace this volume. Does the American mathematician have any comparable device to establish his credentials in Washington?

Armed with Integration... [Vooruzhivshis Integralom...]. By Yu I. Gilderman. Novosibirsk, Nauka, 1980. 192 pp., 35 kopeks.

R. P. BOAS
Department of Mathematics, Northwestern University, Evanston, IL 60201

If you wanted to guess at current trends in education in mathematics in the United States, it would be naive to make inferences from the contents of a single textbook, or even of twenty textbooks. On the other hand, for a country with a centrally controlled educational system, and with the State the only publisher, it may be reasonable to draw conclusions from the kinds of books that appear, especially from those that get sent abroad for review. It seems barely possible that this method is more promising than reading official press releases. It was used by Hilton in the preceding review to estimate the guiding principles of first-year university mathematics in the USSR.

Gilderman's book appears to provide some insight into Soviet high school and college mathematics, more precisely into mathematics outside the elite accelerated schools of mathematics and science. We know that, at least theoretically, many (if not all) Soviet students in what

correspond to the upper high school years in the USA are now exposed to calculus. As our college faculties are only too well aware, the majority of the counterparts of these students in the USA are not being exposed to much of any mathematics at all.

The book is nominally addressed to "inquisitive shkolniky" (ordinarily translated as "schoolchildren") and to students (inquisitive or not) at a somewhat higher level, perhaps corresponding roughly to junior college students. It is a collection of informal essays on applications of mathematics to biology and chemistry, introduced by discussions of the ideas of a mathematical model and of a function. The mathematical topics are enumeration, Boolean algebra, graph theory, linear algebra, and differential equations, so that the book covers in an elementary way a good deal of what Hilton finds missing from the university curriculum. There is a final essay, "Can a machine—have feelings?"

No techniques of calculus are used, and the only idea of calculus that appears is that of the derivative as a rate of change. The students to whom this book is addressed are apparently not acquiring a very deep knowledge of calculus. (The word "integral" is used, not as a mathematical term, but as connoting synthesis or unification.)

One would rather expect that an "inquisitive" high school student in the USA would already have read some of the many books that provide informal presentations of a broad range of mathematical ideas. Is it not legitimate to infer that books on "What is mathematics?" have not met with official approval in the USSR?

In the Soviet system one doesn't just write a book and then try to peddle it to a publisher. Some influential and farsighted person, or group of people, must have commissioned this book, and must have specified that it was to be written in a lively style that would entice students to discover both that there is more mathematics than appears in the curriculum, and that mathematics is even useful. The book culminates in an eloquent peroration:

"People are infinitely complicated but knowledge is always finite. We shall never arrive at the doleful day when everybody knows all about everything. Let us then rejoice in our endless unexplored universe which contains not only rigorous theorems, but also the magical power of words, the inexplicable beauty of sensations, and the elusive secret of imagination. But let us also remember that on the long road from the cave dwellers' campfires to today's space flights, again and again humanity has survived only because it was armed with integration."

I cannot read this as anything other than a plea for a more liberal education and for a recognition of the importance—indeed, the necessity—of a broader view of mathematics. The system described by Hilton is coming under attack from within.

FILMS

CYRIL W. L. GARNER
Department of Mathematics, Carleton University, Ottawa, Ontario, Canada

Dihedral Kaleidoscopes. A film produced by the College Geometry Project at the University of Minnesota. Mathematician: H. S. M. Coxeter. 16 mm sound and color; 13 min. Available for rent or purchase from International Film Bureau, Inc. Sale $195; rental $15. Also available for rent from numerous University Film Libraries.

This is a beautifully crafted film, excelling not only in the ease with which the viewer is introduced to the cyclic and dihedral groups in geometry, but also in the polished perfection of the photography.

Using the concept of bilateral symmetry as it occurs in everyday life, the viewer is introduced to the properties of a plane reflection. This quickly leads to the mathematical abstraction of reflection in a line and rotation as the product of two such reflections in intersecting lines. This can be illustrated most effectively by mirrors, and the photographer has been amazingly successful in capturing the visual effect. The dihedral groups $\mathfrak{D}_1$, $\mathfrak{D}_2$, $\mathfrak{D}_3$, $\mathfrak{D}_4$ are simply presented in this way, and we see why the angle between two mirrors must be a submultiple of π in order to obtain a kaleidoscope instead of chaos; thus the general dihedral group $\mathfrak{D}_n$ of order $2n$, generated by reflections in lines inclined at π/n to each other, is defined. The cyclic subgroups C_n, generated by a rotation through angle $2\pi/n$, are exhibited by a concise treatment of the meaning of sense-preserving and sense-reversing.

The infinite dihedral group is seen to be the limit of $\mathfrak{D}_n$ as the angle between the two mirrors approaches 0, so that the mirrors are parallel. Again by considering "sense," the infinite cyclic group is seen to be a subgroup of index 2 in $\mathfrak{D}_\infty$.

This leads naturally to the third step of difficulty, the study of products of reflections in three mirrors which intersect by pairs in angles which are submultiples of π. For example, the regular tessellation of triangles and the associated rhombic tessellation arise by reflecting a region in the three sides of an equilateral triangle. Similarly the regular pattern of hexagons, the associated semiregular tessellation, and the three ways of obtaining the regular tessellation of squares from an isosceles right triangle are shown. The film ends with a stunning array of colorful tessellations of various degrees of regularity, all obtained by reflecting an appropriate region in three mirrors.

The reviewer has found this film an excellent motivation to the study of symmetry groups, both in algebra and geometry courses; architecture students were especially appreciative of its technical and photographic excellence.

Symmetries of the Cube. A film produced by the College Geometry Project at the University of Minnesota. Mathematician: H. S. M. Coxeter. 16 mm sound and color; 13 1/2 min. Available for rent or purchase from International Film Bureau, Inc. Sale \$195; rental \$15. Also available for rent from numerous University Film Libraries.

As a sequel to "Dihedral Kaleidoscopes," this film shows how the concepts of reflection in a line and rotation about a point can be extended to reflection in a plane and rotation about a line in 3-space. As a particular example, the film exhibits the 9 planes of reflection and the 24 rotations of a cube; it does not mention the other 15 sense-reversing symmetries, the rotatory reflections.

The octahedral kaleidoscope, formed by 3 concurrent plane mirrors inclined at angles of $\pi/2$, $\pi/3$, $\pi/4$ to each other, is introduced after an illustration of sense-preserving and sense-reversing transformations, dihedral groups and cyclic groups. Through skillful photography the viewer is able to see the cube generated by reflecting a suitable object in the mirrors of the kaleidoscope; single framing photography is also used to exhibit the planes of symmetry of the cube and the axes of rotation. This motivates an excellent discussion of the fundamental orthoscheme and its use in the generation of the octahedral group. At the conclusion, other figures which have the same symmetry group are mentioned, notably the octahedron obtained by reciprocating the cube.

As a pedagogical aid, this film would probably be comprehensible only to a student who has previously studied the symmetry group of a cube and has some familiarity with the many terms introduced; it should serve as an incentive to study the symmetry groups of other figures. As a visual experience, the film is a delight, the only unfortunate aspect being the monotone delivery of the narrator which detracts from the excellence of the script.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 9 November 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

BASES IN BANACH SPACES

ROBERT C. JAMES
P. O. Box 1431, Grass Valley, CA 95945

The publication of Banach's book [**2**] in 1932 might be regarded as marking the beginning of the systematic study of Banach spaces. Research activity in this area has expanded dramatically during the past two decades. Interesting new directions have developed and interplays between Banach space theory and other mathematics have proved to be very valuable. Most well-known classical problems have been solved, but some remain and important new problems have arisen. Our purpose is very limited. It is to discuss some of the most important and fundamental facts about bases in Banach spaces, particularly those that may be interesting and useful for mathematicians in other fields. Often, only sketches of proofs will be given, while others are given in more detail because they are relatively easy and may contribute to developing intuitive feeling for the concepts involved. It will be seen that many very important and beautiful theorems have very easy and natural proofs.

A *basis* (or *Schauder basis*) for a Banach space X is a sequence $\{e_n : n \geqslant 1\}$ of members of X which has the property that, for each x in X, there is exactly one sequence of scalars $\{x_i\}$ for which $x = \Sigma_1^\infty x_i e_i$ in the sense that $\lim_{n\to\infty} \|x - \Sigma_1^n x_i e_i\| = 0$. Some important Banach spaces have very natural bases.

If $\{e_n\}$ is a complete orthonormal sequence in Hilbert space H and x is any member of H, then there is exactly one sequence $\{x_n\}$ of scalars such that $\lim_{n\to\infty}\|x - \Sigma_{i=1}^n x_i e_i\| = 0$. For this sequence of scalars, each x_i is (x, e_i) and

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\| = \left[\sum_{i=n+1}^{\infty} |x_i|^2 \right]^{1/2}.$$

Let c_0 be the Banach space of all sequences $x = \{x_i\}$ for which $x_i \to 0$ and $\|x\| = \max\{|x_i| : i \geqslant 1\}$. For each n, let e_n be the member of c_0 that consists of zeros except for 1 in position n. If $x = \{x_i\}$ is any member of c_0, then $\{x_i\}$ is the only sequence of scalars for which $\lim_{n\to\infty}\|x - \Sigma_{i=1}^n x_i e_i\| = 0$. For this sequence of scalars,

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\| = \max\{|x_i| : i > n\}.$$

An important class of Banach spaces is given by the spaces l_p for $1 \leqslant p < \infty$, where l_p is the space of all sequences $x = \{x_i\}$ for which $\Sigma_1^\infty |x_i|^p$ is convergent and $\|x\|$ is $[\Sigma_1^\infty |x_i|^p]^{1/p}$. With $\{e_n\}$ the same as for c_0,

$$\left\| x - \sum_{i=1}^{n} x_i e_i \right\| = \left[\sum_{i=n+1}^{\infty} |x_i|^p \right]^{1/p} \text{ so } \lim_{n\to\infty} \left\| x - \sum_{i=1}^{n} x_i e_i \right\| = 0.$$

After discussing the background that will be useful for understanding what follows, we describe in Section 2 the basis for $C[0,1]$ discovered by Schauder [**26**]. A simple and extremely important characterization of a basis will be introduced in Section 3 and used to show that the "Haar

Robert C. James earned his B.A. at UCLA (1940) and the Ph.D. at Caltech (1947) under the direction of A. D. Michal. He has been at Harvard, U. C. Berkeley, Haverford, Harvey Mudd, and Claremont Graduate School. He also had visiting research appointments at the Institute for Advanced Study (Princeton), the Institute for Advanced Studies (Jerusalem), and the Mittag-Leffler Institute (Sweden). While he was in high school, his family published the *Peace Digest*. Later, his father (Glenn James) and he co-edited the *Mathematics Dictionary*; recent editions have been prepared by him and Ed Beckenbach. Early retirement in 1981 leaves him free to work on research and the building of a new home in Grass Valley, CA (Trees and Bushes, 18040 McCourtney Rd., 95945), a joint project with his long-supportive and capable wife, Edith, and the encouragement of their four married children.

system" is a basis for $L_p[0, 1]$ if $1 \leqslant p < \infty$. In Section 4, we show that any sequence "sufficiently close" to a basis is itself a basis. In Sections 5 and 6, duality and reflexivity are related to the concepts of bases that are "shrinking" and/or "boundedly complete," and these ideas are used to understand the nonreflexive Banach space J that is isometric with its second dual. In Sections 7 and 8, we will discuss many implications of the existence of an unconditional basis, and see that some common spaces do not have unconditional bases. We conclude with a discussion of some solved and unsolved problems in Section 9.

1. Background. A *Banach space* is a linear space X with either real or complex scalars for which each x in X has a norm $\|x\|$ for which $\|x\| > 0$ if $x \neq 0$, $\|x + y\| \leqslant \|x\| + \|y\|$ for all x and y in X (the *triangle inequality*), $\|ax\| = |a|\,\|x\|$ for all scalars a and members x of X, and X is a complete metric space with respect to the distance defined by $d(x, y) = \|x - y\|$. Hereafter, *space* always will mean *Banach space* and the symbols X and Y will be used only to denote Banach spaces. We will use $\mathrm{lin}(A)$ to denote the algebraic linear span of the set A; *i.e.*, the set of all finite linear combinations of members of A. The closure of $\mathrm{lin}(A)$ will be denoted by $\mathrm{cl}[\mathrm{lin}(A)]$. A Banach space is *separable* if it contains a dense sequence; this is equivalent to containing a sequence $\{x_n\}$ for which X is $\mathrm{cl}[\mathrm{lin}\{x_n\}]$.

A linear mapping T of X into Y is continuous if and only if it is continuous at 0, or if and only if there is a number $\|T\|$ for which

$$\sup\{\|Tx\| : \|x\| \leqslant 1\} = \|T\| < \infty. \tag{1}$$

Also, $\|T\|$ is the least number M such that $\|Tx\| \leqslant M\|x\|$ for each x. Two spaces X and Y are *isomorphic* if there is an algebraic isomorphism T of X onto Y for which both T and T^{-1} are continuous. In this case, there are positive numbers α and β for which

$$\alpha\|x\| \leqslant \|Tx\| \leqslant \beta\|x\| \quad \text{if } x \in X.$$

If $\alpha = \beta = 1$, then X and Y are said to be *isometric*. If Y is simply X with a new norm, then the old and new norms are *equivalent* if X and Y are isomorphic with T the identity map.

A *linear functional* on X is a continuous linear mapping from X into the space of scalars. The linear space of all linear functionals on X is complete with respect to the norm given by (1). It is the *first dual* of X, and is denoted by X^*. If ϕ is a linear functional on X, then $\phi(x)$ usually will be denoted by (ϕ, x) or (x, ϕ).

It would be helpful if the reader is familiar with the preceding and with other material that now is a part of most standard texts in real analysis and measure theory, especially the definition of inner-product spaces, properties of orthonormal bases, and the two following theorems. *Hilbert space* will mean a separable infinite-dimensional inner-product space.

HAHN - BANACH THEOREM. *If ϕ is a linear functional on a linear subset L of a Banach space X, then there is a linear functional Φ on X for which $\|\phi\|_L = \|\Phi\|$ and $(\Phi, x) = (\phi, x)$ if $x \in L$.*

Rather than making explicit use of the Hahn-Banach theorem, one often uses the consequence that, for each nonzero x in X, there is a linear functional x^* for which $\|x^*\| = 1$ and $(x^*, x) = \|x\|$. Usually we will not distinguish between the theorem and this easy consequence of it.

INVERSE - MAPPING THEOREM (*Open-mapping theorem for one-to-one mappings*). *If T is a continuous linear one-to-one mapping of a Banach space X onto a Banach space Y, then T^{-1} is continuous.*

Any complex space becomes a real space if one agrees to use only real scalars. Also, each complex linear functional f determines two real linear functionals, $\theta = \mathrm{Re}(f)$ and $\phi = \mathrm{Im}(f)$, for which $\phi(x) = -\theta(ix)$ and $\|\theta\| = \|\phi\| = \|f\|$. Any real linear functional is the real part of exactly one complex linear functional. Therefore, in a very real sense, the theory of complex Banach spaces is less general than the theory of real Banach spaces. In fact, a complex Banach space is simply a real Banach space for which "many" two-dimensional subspaces are Euclidean.

Therefore, we will not hesitate to use real spaces for examples or to restrict a discussion to real spaces. However, in many cases it will be immaterial whether the scalars are real or complex and we will not specify which is intended.

2. The Space $C[0,1]$. If all separable Banach spaces had bases, especially if a basis always could be found as easily as for H, c_0, and the l_p-spaces, then the subject would be much simpler. The space $C[0,1]$ provides a simple example for which a basis is not so obvious; $C[0,1]$ is the real space of real-valued functions that are continuous on the closed interval $[0,1]$, with $\|f\|$ the maximum of $|f(x)|$ on $[0,1]$. With this norm, convergence of a sequence of continuous functions in $C[0,1]$ means uniform convergence. To describe a basis $\{f_n\}$ for $C[0,1]$, let $\{t_i\colon i \geqslant 1\}$ be the sequence of dyadic points in $[0,1]$: $0, 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \frac{3}{16}, \ldots$. Let f_1 and f_2 be identically 1 and t, respectively. For each $n > 2$, let $f_n(t_j) = 0$ if $j < n$, $f_n(t_n) = 1$, and f_n be linear between any two neighbors among the first n dyadic points. To show that $\{f_n\}$ is a basis for $C[0,1]$, we will show that each member g of $C[0,1]$ has a unique representation as $\Sigma_1^\infty a_i f_i$. Note first that a_1 must be $g(0)$, since $f_i(0) = 0$ if $i > 1$. Since $g - a_1 f_1 = \Sigma_2^\infty a_i f_i$ and $f_i(1) = 0$ if $i > 2$, a_2 must be $g(1) - a_1 f_1(1) = g(1) - a_1$. This can be continued inductively to determine all coefficients uniquely, with

$$a_n = g(t_n) - \sum_{i=1}^{n-1} a_i[f_i(t_n)].$$

Then $p_n = \Sigma_1^n a_i f_i$ is a polygonal function with endpoints and vertices on the graph of g at points whose abscissas are the first n dyadic points, $t_1, t_2, \ldots, t_n$. Since $\{t_i\}$ is dense in $[0,1]$ and g is uniformly continuous, $\{p_n\}$ converges uniformly to g. That is,

$$\lim_{n\to\infty}\left\| g - \sum_{i=1}^{n} a_i f_i \right\| = 0 \quad \text{and} \quad g = \sum_{i=1}^{\infty} a_i f_i .$$

The basis we have obtained for $C[0,1]$ is *normalized*; that is, $\|f_n\|$ is 1 for each n.

3. Bases and Projections. For a sequence $\{e_n\}$ to be a basis for X, it clearly is necessary for the linear span of $\{e_n\}$ to be dense in X. This is not sufficient, even with the assumption of linear independence. For example, the sequence of polynomials $\{t^{n-1} : n \geqslant 1\}$ is not a basis for $C[0,1]$, since no continuous function that is not differentiable at 0 has a power series representation that converges uniformly on $[0,1]$—or even converges for some nonzero t.

The following extremely useful theorem describes a kind of "independence" that is sufficient. Before proving this theorem, let us establish an inequality (2) that often will be useful. Suppose the linear span of $\{e_n\}$ is dense in X, no e_n is 0, and there is a positive K such that, for all positive integers n and p and scalars $\{a_i\}$,

$$K\left\| \sum_{i=1}^{n+p} a_i e_i \right\| \geqslant \left\| \sum_{i=1}^{n} a_i e_i \right\| .$$

For each k, let x_k^* be a linear functional defined on $\operatorname{lin}\{e_n\}$ by letting $(e_k^*, \Sigma_1^n a_i e_i)$ be a_k if $n \geqslant k$ and 0 otherwise. Then each e_k^* is continuous. In fact, for any $x = \Sigma_1^\infty x_i e_i$ with finitely many nonzero terms,

$$|(e_k^*, x)| = |x_k| = \left\| \sum_1^k x_i e_i - \sum_1^{k-1} x_i e_i \right\| / \|e_k\| \leqslant 2K\|x\|/\|e_k\|,$$

so

$$\|e_k^*\| \leqslant \frac{2K}{\|e_k\|}, \tag{2}$$

and e_k^* can be extended by continuity to all of X. These linear functionals $\{e_n^*\}$ are the *coefficient functionals* of $\{e_n\}$.

THEOREM 3.1. *If the linear span of $\{e_n\}$ is dense in X and no e_n is 0, then $\{e_n\}$ is a basis for X if and only if there is a positive K such that, for all positive integers n and p and scalars $\{a_i\}$,*

$$K\left\|\sum_{i=1}^{n+p} a_i e_i\right\| \geqslant \left\|\sum_{i=1}^{n} a_i e_i\right\|. \tag{3}$$

The least number K that satisfies (3) is the *basis constant* of $\{e_n\}$. If the basis constant is 1, then $\|\Sigma_1^n a_i e_i\|$ is a monotone increasing function of n and the basis is said to be *monotone*. It is customary to say that x is *orthogonal* to y if $\|x + ky\| \geqslant \|x\|$ for all scalars k [**29**(I), p. 215]. Thus for a monotone basis, lin$\{e_i$: $1 \leqslant i \leqslant n\}$ is orthogonal to cl[lin$\{e_i$: $i > n\}$]. Any basis $\{e_n\}$ becomes monotone if the norm is replaced by the norm $|||\ \ |||$ for which

$$|||\,x\,||| = \sup\left\{\left\|\sum_{i=1}^{n} x_i e_i\right\|:\ n \geqslant 1\right\},\quad \text{if } x = \sum_{i=1}^{\infty} x_i e_i. \tag{4}$$

The proof of Theorem 3.1 will be sketched briefly (also see [**2**, p. 111], [**23**(I), p. 2]). To prove the necessity, we first introduce the new norm given by (4). It can be shown that X is complete if given this norm. The identity mapping T from X with $|||\ \ |||$ to X with $\|\ \ \|$ is continuous, so it follows from the inverse-mapping theorem that T^{-1} is continuous. This gives the existence of K. To prove sufficiency, one first shows that each x in X can be represented as

$$x = \sum_{i=1}^{n} (e_i^*, x) e_i + h_n,$$

where h_n belongs to cl[lin$\{e_i$: $i > n\}$]. Now for x and an arbitrary $\Delta > 0$, choose N and numbers $\{c_i\}$ for which $c_i = 0$ if $i > N$ and $\|x - \Sigma_1^N c_i e_i\| < \Delta$. Then

$$\left\|\sum_{i=1}^{n} [(e_i^*, x) - c_i] e_i + h_n\right\| < \Delta \quad \text{if} \quad n \geqslant N.$$

This and (3) imply $K\Delta > \|\Sigma_1^n[(e_i^*, x) - c_i]e_i\|$, and with use of the triangle inequality we have $\|h_n\| < \Delta(1 + K)$.

Theorem 3.1 often is described in terms of projections. If $\{e_i\}$ is a basis for X and we define $P_n(x)$ to be $\Sigma_1^n x_i e_i$ if $x = \Sigma_1^\infty x_i e_i$, then $P_n^2 = P_n$. Also, (3) implies $K\|x\| \geqslant \|P_n(x)\|$, so P_n is continuous and $\|P_n\| \leqslant K$. Thus all such P_n are projections and their norms are bounded by K. The "if and only if" condition in Theorem 3.1 could be replaced by "there is a sequence $\{P_n\}$ of uniformly bounded projections on X for which the range of P_n is lin$\{e_i$: $i \leqslant n\}$ and $P_n(e_i) = 0$ if $i > n$."

The most natural use for Theorem 3.1 is for proving that a candidate for a basis actually is a basis. We established the basis $\{f_n\}$ for $C[0, 1]$ rather easily. However, it is interesting to note that K of Theorem 3.1 is 1 for this sequence, since $|\Sigma_1^n a_i f_i(t)|$ has its largest value at some t_k with $k \leqslant n$, but $\Sigma_1^{n+p} a_i f_i(t_k) = \Sigma_1^n a_i f_i(t_k)$ if $k \leqslant n$. There are many bases for which the use of Theorem 3.1 is more vital. For example, let $L_p[0, 1]$ for $1 \leqslant p < \infty$ be the Banach space of all Lebesgue-measurable real-valued functions on $[0, 1]$, with

$$\|f\| = \left[\int_{[0,1]} |f|^p\right]^{1/p} < \infty.$$

As a candidate for a basis, we choose the *Haar system* $\{\phi_i\}$ defined as follows (see [**23**], [**24**], [**29**]). The function ϕ_1 is identically 1 on $[0, 1]$; ϕ_2 is 1 on $[0, \frac{1}{2})$ and -1 on $[\frac{1}{2}, 1]$; ϕ_3 is 1 on $[0, \frac{1}{4})$, -1 on $[\frac{1}{4}, \frac{1}{2})$, and 0 on $[\frac{1}{2}, 1]$; ϕ_4 is 1 on $[\frac{1}{2}, \frac{3}{4})$, -1 on $[\frac{3}{4}, 1]$, and 0 on $[0, \frac{1}{2})$. For positive integers r and $k \leqslant 2^{r-1}$, $\phi_{2^{r-1}+k}$ is 1 on I_{2k-1}^r, -1 on I_{2k}^r, and 0 otherwise, where $[0, 1]$ is partitioned into 2^r

intervals $\{I_j^r\colon 1 \leqslant j \leqslant 2^r\}$ of equal lengths. Except for endpoints, $\mathrm{lin}\{\phi_i\}$ contains all characteristic functions of these intervals $\{I_j^r\}$ and therefore is dense in $L_p[0,1]$. If $f = \Sigma_1^n a_i\phi_i$ and $g = \Sigma_1^{n+1} a_i\phi_i$, then $f(t) = g(t)$ except on some interval on which f has a constant value α and g is constantly $\alpha + a_{n+1}$ on the first half and constantly $\alpha - a_{n+1}$ on the second half. Since $|t|^p$ is convex if $p \geqslant 1$,

$$|\alpha + a_{n+1}|^p + |\alpha - a_{n+1}|^p \geqslant 2|\alpha|^p,$$

and therefore $\|g\| \geqslant \|f\|$. It now follows that $\{\phi_i\}$ is a basis and the basis constant is 1. This basis for $L_p[0,1]$ has long been known [**27**]. The usual trigonometric (Fourier) system is a basis for $L_p[0,1]$ if $1 < p < \infty$, but not if $p = 1$ [**24**, pp. 51-53]. If each of the functions in the basis $\{\phi_i\}$ for $L_p[0,1]$ is divided by its norm in $L_2[0,1]$, we obtain the sequence of *Haar functions*, which is an orthonormal basis for $L_2[0,1]$.

4. Equivalent bases. *Equivalent bases* for X are bases $\{u_n\}$ and $\{v_n\}$ for which $\Sigma_1^\infty a_n u_n$ converges if and only if $\Sigma_1^\infty a_n v_n$ converges. If X is infinite-dimensional and has a basis, then there are uncountably many bases for X, no two of which are equivalent [**23**(I), p. 5]. However, bases do have some stability. If each basis vector is perturbed by a sufficiently small amount, it remains a basis and is equivalent to the original basis.

THEOREM 4.1. *If $\{u_n\}$ is a normalized basis for X and $\{u_n^*\}$ is the corresponding sequence of coefficient functionals, then $\{v_n\}$ is a basis for X and is equivalent to $\{u_n\}$ if*

$$\sum_{i=1}^{\infty} \|u_i - v_i\|\,\|u_i^*\| < 1. \tag{5}$$

Proof. Let K be the basis constant of $\{u_n\}$ and let θ denote the left member of (5). We define a linear map T from X into the linear space of formal sums $\Sigma_1^\infty a_i v_i$ by letting $Tx = \Sigma_1^\infty x_i v_i$ if $x = \Sigma_1^\infty x_i u_i$. Since $\{u_n\}$ is normalized, it follows from (2) that $\|u_n^*\| \leqslant 2K$ for each n. Then

$$|x_n| = |(u_n^*, x)| \leqslant \|u_n^*\|\,\|x\|,$$

so

$$\|x - Tx\| = \left\| \sum_{i=1}^{\infty} x_i(u_i - v_i) \right\| \leqslant \|x\| \sum_{i=1}^{\infty} \|u_i - v_i\|\,\|u_i^*\| = \theta\|x\|$$

and $\Sigma_1^\infty x_i v_i$ converges. Also,

$$\|x\| - \|x - Tx\| \leqslant \|Tx\| \leqslant \|x\| + \|x - Tx\|,$$

so

$$(1 - \theta)\|x\| \leqslant \|Tx\| \leqslant (1 + \theta)\|x\|$$

and T is an isomorphism. Since $Tu_n = v_n$ for each n, no v_n is 0. Since $\|I - T\| \leqslant \theta < 1$, the range of T is all of X and therefore $\mathrm{lin}\{v_i\}$ is dense in X. Also, if $\sigma_n = \Sigma_1^n a_i u_i$ and $\sigma_{n+1} = \Sigma_1^{n+p} a_i u_i$, then

$$\|T\sigma_n\| \leqslant (1 + \theta)\|\sigma_n\| \leqslant (1 + \theta)K\|\sigma_{n+1}\| \leqslant \frac{1 + \theta}{1 - \theta} K\|T\sigma_{n+1}\|,$$

so it follows from Theorem 3.1 that $\{v_n\}$ is a basis for X with basis constant $K(1 + \theta)/(1 - \theta)$. We know that $\Sigma_1^\infty a_i v_i$ converges if $\Sigma_1^\infty a_i u_i$ converges. It now follows that $\Sigma_1^\infty a_i u_i$ converges if $\Sigma_1^\infty a_i v_i$ converges, since each member of X has exactly one representation of type $\Sigma_1^\infty a_i v_i$. This completes the proof of Theorem 4.1. Often (5) is replaced by the stronger condition $\Sigma_1^\infty \|u_i - v_i\| < 1/(2K)$.

The basis we established for $C[0,1]$ was a sequence of polygonal functions. Because of Theorem 4.1, we know that $C[0,1]$ has a basis of polynomials. However, the degrees of the polynomials increase rapidly. Similarly, the Haar system for $L_p[0,1]$ could be replaced (for each $1 \leqslant p < \infty$) by a basis of continuous functions or a basis of polynomials.

5. Bases and Duality. With the exception of the second half of Theorem 5.3, the theorems of this and the next section come from [**15**] and [**16**].

Let $\{e_n\}$ be the natural basis for the space c_0. If y is a linear functional on c_0, then for each $x = \Sigma_1^\infty x_i e_i$ we have

$$(y, x) = \sum_{i=1}^{n} x_i y_i + \left(y, \sum_{i=n+1}^{\infty} x_i e_i\right),$$

where $y_i = (y, e_i)$ for each i. Since $\|\Sigma_{n+1}^\infty x_i e_i\| \to 0$, this implies $(y, x) = \Sigma_1^\infty x_i y_i$ and

$$|(y, x)| = \left|\sum_{i=1}^{\infty} x_i y_i\right| \leqslant \max\{|x_i|\} \sum_{i=1}^{\infty} |y_i| \quad \text{for all } x. \tag{6}$$

Thus $\|y\| \leqslant \Sigma_1^\infty |y_i|$. By letting x_i be sign(y_i) for many values of i, and 0 thereafter, we can see that $\Sigma_1^\infty |y_i|$ cannot be replaced in (6) by a smaller number. This implies $\|y\| = \Sigma_1^\infty |y_i|$ and that the dual of c_0 is l_1. Since $y = \Sigma_1^\infty y_i e_i^*$ if $\{e_n^*\}$ is the sequence of coefficient functionals for $\{e_n\}$, we also know that $\{e_n^*\}$ is the natural basis for l_1. This might lead one to wonder whether there are other spaces X with a basis $\{e_n\}$ for which $\{e_n^*\}$ is a basis for X^*. It is easy to see that this is not always the case. By an argument similar to the preceding, one can show that the dual of l_1 is the space l_∞ of all bounded sequences $x = \{x_i\}$ with $\|x\| = \sup\{|x_i|\}$. But this space is not separable and therefore cannot have a basis.

However, it is true that $\{e_n^*\}$ always is a *basic sequence*; *i.e.*, it is a basis for cl[lin$\{e_n^*\}$]. To show this, let n and p be positive integers, $\{a_i\}$ an arbitrary sequence of scalars, and δ an arbitrary positive number. Then use the definition of $\|\Sigma_1^n a_i e_i^*\|$ to get an $x = \Sigma_1^\infty c_i e_i$ for which $\|x\| = 1$ and

$$\left(\sum_1^n a_i e_i^*, x\right) > \left\|\sum_1^n a_i e_i^*\right\| - \delta.$$

If K is the basis constant of $\{e_n\}$, then

$$K = K\|x\| \geqslant \left\|\sum_1^n c_i e_i\right\|$$

and therefore

$$K\left\|\sum_{i=1}^{n+p} a_i e_i^*\right\| \geqslant \left\|\sum_{i=1}^{n+p} a_i e_i^*\right\| \left\|\sum_{i=1}^{n} c_i e_i\right\| \geqslant \left(\sum_{i=1}^{n+p} a_i e_i^*, \sum_{i=1}^{n} c_i e_i\right)$$
$$= \sum_{i=1}^{n} a_i c_i = \left(\sum_{i=1}^{n} a_i e_i^*, x\right) > \left\|\sum_{i=1}^{n} a_i e_i^*\right\| - \delta.$$

Since δ was arbitrary,

$$K\left\|\sum_1^{n+p} a_i e_i^*\right\| \geqslant \left\|\sum_1^n a_i e_i^*\right\|.$$

That is, $\{e_i^*\}$ is a basic sequence with basis constant not greater than K.

We have seen that $\{e_n^*\}$ is a basic sequence, so it is a basis for X^* if and only if lin$\{e_n^*\}$ is dense in X^*. The next theorem gives a very interesting and useful test for this.

THEOREM 5.1. *Let $\{e_n\}$ be a basis for X and let $\{e_n^*\}$ be the coefficient functionals. Then each of the following is a necessary and sufficient condition for $\{e_n^*\}$ to be a basis for X^*.*

(i) cl[lin$\{e_n^*\}$] $= X^*$.
(ii) $\lim_{n\to\infty} \|x^*\|_n = 0$ *for each x^* in X^*, where $\|x^*\|_n$ is the norm of x^* when x^* is restricted to* lin$\{e_i: i > n\}$.

A basis which has property (ii) is said to be *shrinking*. The natural basis $\{e_n\}$ of c_0 is shrinking. Each x^* in the dual l_1 of c_0 is representable as $\Sigma_1^\infty a_i e_i^*$, and

$$\left\|x^* - \sum_{i=1}^{n} a_i e_i^*\right\| = \|x^*\|_n = \sum_{i=n+1}^{\infty} |a_i|,$$

which approaches 0 as n increases.

Any orthogonal basis for Hilbert space is shrinking. More generally, the natural basis of l_p is shrinking if $1 < p < \infty$. To see this, suppose there is a linear functional ϕ defined on l_p that does not "shrink"; *i.e.*, $\lim_{n\to\infty}\|\phi\|_n = \beta > 0$. Then for any m and for a suitable n greater than m, there are members u and v of l_p for which u belongs to $\text{lin}\{e_i: m \leqslant i \leqslant n\}$, v belongs to $\text{lin}\{e_i: i > n\}$, $\|u\| = \|v\| = 1$, and both (ϕ, u) and (ϕ, v) are nearly as great as β. Then $(\phi, u + v)$ is approximately 2β, so if m had been chosen great enough that $\|\phi\|_m$ is approximately equal to β, then $\|u + v\|$ is approximately equal to 2. We have the contradiction that 2 is approximately equal to $\|u + v\|$, which is equal to

$$[\|u\|^p + \|v\|^p]^{1/p} = 2^{1/p}.$$

The natural basis $\{e_n\}$ of l_1 is not shrinking, since there is a member x^* of the dual l_∞ for which $(x^*, e_n) = 1$ for all n and $\|x^*\|_n = 1$ for all n. The functional x^* does not "shrink" at all.

The proof of Theorem 5.1 can be very short. We have seen already that (i) is necessary and sufficient. Now consider (ii), and suppose $\{e_n\}$ is shrinking and the basis constant is K. If x^* belongs to X^*, then the linear functional $x^* - \Sigma_1^n(x^*, e_i)e_i^*$ is identically 0 on $\text{lin}\{e_i: i \leqslant n\}$. Thus if $x = \Sigma_1^\infty x_i e_i$ and $\|x\| = 1$, then for each n we have

$$\left\|\sum_{i=n+1}^{\infty} x_i e_i\right\| = \left\|x - \sum_{i=1}^{n} x_i e_i\right\| \leqslant \|x\| + \left\|\sum_{i=1}^{n} x_i e_i\right\| \leqslant 1 + K,$$

and

$$\left|\left(x^* - \sum_{i=1}^{n} x^*(e_i)e_i^*, x\right)\right| = \left|\left(x^*, \sum_{i=n+1}^{\infty} x_i e_i\right)\right| \leqslant \|x^*\|_n \left\|\sum_{i=n+1}^{\infty} x_i e_i\right\|$$
$$\leqslant (1 + K)\|x^*\|_n,$$

which approaches 0 as n increases. Therefore, $\|x^* - \Sigma_1^n(x^*, e_i)e_i^*\|$ approaches 0 and

$$x^* = \sum_{1}^{\infty}(x^*, e_i)e_i^*,$$

so $\{e_n^*\}$ is a basis for X^*. The converse is easier. If $\{e_n^*\}$ is a basis for X^* and $x^* = \Sigma_1^\infty \xi_i e_i^*$, then

$$\|x^*\|_n = \left\|\sum_{n+1}^{\infty} \xi_i e_i^*\right\|$$

and this approaches 0 as n increases.

DEFINITION 5.2. A *boundedly complete* basis for a Banach space X is a basis $\{e_n\}$ for which $\Sigma_1^\infty a_i e_i$ converges whenever the sequence of scalars $\{a_n\}$ has the property that $\|\Sigma_1^n a_i e_i\|$ is a bounded function of n.

The natural basis $\{e_n\}$ of c_0 is not boundedly complete, since $\|\Sigma_1^n e_i\| = 1$ for all n, but $\Sigma_1^\infty e_i$ is not convergent.

The natural basis $\{e_n\}$ of l_1 is boundedly complete, since if $\|\Sigma_1^n a_i e_i\| = \Sigma_1^n |a_i|$ is a bounded function of n, then $\Sigma_1^\infty |a_i| < \infty$.

The natural basis $\{e_n\}$ of l_p $(1 \leqslant p < \infty)$ is boundedly complete, since if $\{a_n\}$ is a sequence of scalars for which

$$\left\|\sum_1^n a_i e_i\right\| = \left[\sum_1^n |a_i|^p\right]^{1/p}$$

is bounded, then $\Sigma_1^\infty |a_i|^p < \infty$ and $\Sigma_1^\infty a_i e_i$ is a member of the space. In particular, any orthogonal basis for Hilbert space is boundedly complete.

The next theorem shows that "boundedly complete" is in some senses dual to "shrinking."

THEOREM 5.3. *If a basis $\{e_n\}$ for a Banach space X is shrinking, then the basis $\{e_n^*\}$ for X^* is boundedly complete. If a basis $\{e_n\}$ for X is boundedly complete, then X is isomorphic to the dual of a Banach space that has a shrinking basis.*

To see that the first statement of this theorem is true, we recall first that, if $\{e_n\}$ is shrinking, then $\{e_n^*\}$ is a basis for X^*. To show that $\{e_n^*\}$ is boundedly complete, suppose $\{a_n\}$ is a sequence of scalars for which there is a number M such that $\|\Sigma_1^n a_i e_i^*\| \leqslant M$ for all $n \geqslant 1$. Suppose we could define a linear functional x^* by letting (x^*, x) be $\Sigma_1^\infty a_i(e_i^*, x)$ for each x in X. Then we would have

$$\left|\sum_1^\infty a_i(e_i^*, x)\right| = \lim_{n\to\infty}\left|\left(\sum_1^n a_i e_i^*, x\right)\right| \leqslant M\|x\|,$$

so $\|x^*\| \leqslant M$, $x^* \in X^*$, and $x^* = \Sigma_1^\infty a_i e_i^*$. Therefore we need only show that $\lim_{n\to\infty}\Sigma_1^n a_i(e_i^*, x)$ exists for each x in X. We do this by showing that the Cauchy convergence condition is satisfied:

$$\left|\sum_{i=m}^n a_i(e_i^*, x)\right| = \left|\left(\sum_{i=m}^n a_i e_i^*, x\right)\right| = \left|\left(\sum_{i=m}^n a_i e_i^*, \sum_{i=m}^n (e_i^*, x)e_i\right)\right|$$
$$\leqslant \left\|\sum_{i=m}^n a_i e_i^*\right\| \left\|\sum_{i=m}^n (e_i^*, x)e_i\right\| \leqslant 2M\left\|\sum_{i=m}^n (e_i^*, x)e_i\right\|,$$

which approaches 0 as m increases with $m \leqslant n$.

The proof of the second part of Theorem 5.3 also is not difficult, but we will let the reader fill in the details (see [**23**(I), p. 9]). One recalls first that $\{e_n^*\}$ is a basis for $Y = \mathrm{cl}[\mathrm{lin}\{e_n^*\}]$. Then the fact that $\{e_n\}$ is boundedly complete can be used to show that Y^* is isomorphic to X and that $\{e_n^*\}$ is a shrinking basis for Y.

We have seen that the dual X^* of X has a basis if X has a shrinking basis, but X^* can be nonseparable and not have a basis even if X has a basis. This problem does not arise in the other "direction": If X^* has a basis, then X has a basis. In fact, X has a shrinking basis and therefore X^* has a boundedly complete basis [**20**, Theorem 1.2].

We may not know much more about X^* other than that it has a boundedly complete basis, if all we know about X is that it has a shrinking basis. However, it is possible to give a very useful description of X^{**}.

THEOREM 5.4. *Let $\{e_n\}$ be a shrinking basis for X. Then the correspondence*

$$x^{**} \leftrightarrow \{(x^{**}, e_1^*), (x^{**}, e_2^*), (x^{**}, e_3^*), \ldots\} \tag{7}$$

is an algebraic isomorphism of X^{**} with the space of all sequences of scalars $\{a_n\}$ such that $\|\Sigma_1^n a_i e_i\|$ is a bounded function of n. If $\{e_n\}$ is monotone, then

$$\|x^{**}\| = \lim_{n\to\infty}\left\|\sum_{i=1}^n (x^{**}, e_i^*)e_i\right\|. \tag{8}$$

Proof. There is no loss of generality if we assume $\{e_n\}$ to be monotone, in which case the limit in (8) exists. Let us show first that the linear mapping defined by (7) is one-to-one and that (8) is satisfied. Since $\{e_n\}$ is shrinking, $\{e_n^*\}$ is a basis for X^* and any x^{**} is determined by its values on

$\{e_n^*\}$. Thus the linear mapping in (7) is one-to-one. To verify (8), suppose $x^{**} \in X^{**}$ and that $x^* = \sum_1^\infty \xi_i e_i^*$ is an arbitrary member of X^*. Then

$$|(x^{**}, x^*)| = \lim_{n\to\infty}\left|\sum_{i=1}^n (x^{**}, e_i^*)\xi_i\right| = \lim_{n\to\infty}\left|\left(x^*, \sum_{i=1}^n (x^{**}, e_i^*)e_i\right)\right|$$
$$\leqslant \lim_{n\to\infty}\|x^*\|\left\|\sum_{i=1}^n (x^{**}, e_i^*)e_i\right\|, \tag{9}$$

so $\|x^{**}\| \leqslant \lim_{n\to\infty}\|\sum_1^n(x^{**}, e_i^*)e_i\|$. For any n, the definition of the norm of a linear functional implies that x^* in (9) could have been chosen so that $\|x^*\| = 1$ and $(x^*, \sum_1^n(x^{**}, e_i^*)e_i)$ is approximately equal to $\|\sum_1^n(x^{**}, e_i^*)e_i\|$. Therefore,

$$\|x^{**}\| \geqslant \left\|\sum_1^n (x^{**}, e_i^*)e_i\right\|$$

for each n and $\|x^{**}\|$ is given by (8). It remains to show that a sequence of scalars $\{a_n\}$ corresponds to some x^{**} if $\|\sum_1^n a_i e_i\|$ is a bounded function of n. Suppose $\|\sum_1^n a_i e_i\| \leqslant M$ for all n. Then we can define x^{**} by letting

$$(x^{**}, x^*) = \lim_{n\to\infty}\left(x^*, \sum_1^n a_i e_i\right) \quad \text{if} \quad x^* \in X^*, \tag{10}$$

which converges because $\{e_n\}$ is shrinking. Since $|(x^{**}, x^*)| \leqslant \|x^*\|M$, we have $\|x^{**}\| \leqslant M$ and therefore $x^{**} \in X^{**}$. It also follows from (10) that $(x^{**}, e_i^*) = a_i$ for each i, so $x^{**} \leftrightarrow \{a_i\}$.

The original purpose of the preceding theorem was to lay the foundations for the Banach space J [**15**, (iii) p. 525].

Example 5.5. The Banach space J consists of all sequences of real numbers $x = \{x_n\}$ for which $\lim_{n\to\infty} x_n$ is 0 and $\|x\| < \infty$, where

$$\|x\| = \sup\left\{\left[(x_{p_1} - x_{p_2})^2 + (x_{p_2} - x_{p_3})^2 + \cdots + (x_{p_{n-1}} - x_{p_n})^2 + (x_{p_n} - x_{p_1})^2\right]^{1/2}\right\} \tag{11}$$

and the supremum is taken over all positive integers n and all increasing sequences of positive integers $\{p_1, p_2, \ldots, p_n\}$ (see [**16**], [**23**(I), p. 25]).

The natural basis of J is $\{e_n\}$, where $\{e_n\}$ is the sequence of all zeros except for 1 in position n. This basis is monotone and shrinking. By use of Theorem 5.4, one can show easily that J^{**} can be described exactly as J, except for dropping the restriction that $\lim_{n\to\infty} x_n = 0$. If J is the set of all sequences in J^{**} that are also in J, then J^{**} is the linear span of J and the sequence $\{1, 1, 1, \ldots\}$ with all terms 1. Thus J is isometric to a maximal closed proper subspace of J^{**}. However, if T is defined by

$$T(x_1, x_2, \ldots) = (x_2 - x_1, x_3 - x_1, \ldots, x_n - x_1, \ldots),$$

then T is an isometric map of J onto J^{**} [**16**]. The term $(x_{p_n} - x_{p_1})^2$ in (11) was introduced only so that T would be an isometry rather than merely an isomorphism.

The space J was used to disprove several long-standing conjectures: a Banach space X is reflexive if X^{**} is separable (or if X^{**} is isomorphic to X); any infinite-dimensional real Banach space is isomorphic to the real space obtained from some complex Banach space by using only real scalars [**11**] (this is true for finite-dimensional Banach spaces if and only if the dimension is even); any infinite-dimensional Banach space X is isomorphic to $X \times X$ [**5**]. Many other applications of J have been found (*e.g.*, see [**1**], [**3**], [**7**], [**18**], [**22**], [**23**(I), p. 25, 103, 132], [**23**(II), p. 36, 39]).

6. Reflexive Banach Spaces. Any Banach space X has a natural isometric mapping T into its second dual X^{**}, for which Tx is the member of X^{**} defined by

$$(Tx, x^*) = (x^*, x) \quad \text{if } x^* \in X^*. \tag{12}$$

To see that $\|Tx\| = \|x\|$, one first observes that

$$|(Tx, x^*)| = |(x^*, x)| \leqslant \|x^*\| \, \|x\| \quad \text{for each } x^*, \tag{13}$$

which implies $\|Tx\| \leqslant \|x\|$. However, because of the Hahn-Banach theorem, we could choose x^* so that the inequality in (13) is arbitrarily close to being an equality. Therefore $\|Tx\| = \|x\|$.

DEFINITION 6.1. A *reflexive* Banach space is a Banach space X for which the natural mapping (12) of X into X^{**} has X^{**} as its range.

For Theorem 5.4, we assumed that the basis $\{e_n\}$ of X is shrinking. Suppose $\{e_n\}$ also is boundedly complete. Then for the sequence $\{a_n\}$ used in (10) for which $\|\Sigma_1^n a_i e_i\|$ is bounded, the series $\Sigma_1^\infty a_i e_i$ is convergent and x^{**} as defined by (10) is the natural image in X^{**} of $\Sigma_1^\infty a_i e_i$ in X. This gives the "if" part of the following very useful characterization of reflexivity in terms of bases [**15**, Theorem 1].

THEOREM 6.2. *If X has a basis $\{e_n\}$, then X is reflexive if and only if $\{e_n\}$ is both shrinking and boundedly complete.*

To complete the proof of this theorem, we must verify the "only if" part. Suppose first that X is reflexive and has a basis $\{e_n\}$ for which cl[lin$\{e_n^*\}$] is not all of X^*. Then it follows from the Hahn-Banach theorem that there is a nonzero member x^{**} of X^{**} that is 0 on cl[lin$\{e_n^*\}$]. But since X is reflexive, there is a member $\Sigma_1^\infty a_i e_i$ of X for which

$$(x^{**}, x^*) = \left(x^*, \sum_1^\infty a_i e_i\right) \quad \text{if } x^* \in X^*.$$

This is not possible, since $(x^{**}, e_n^*) = 0 = (e_n^*, \Sigma_1^\infty a_i e_i) = a_n$ for each n. Therefore $\{e_n^*\}$ spans X^*, which we have seen implies $\{e_n\}$ is shrinking. It remains to show that $\{e_n\}$ is boundedly complete. Observe that the natural image in X^{**} of any member $x = \Sigma_1^\infty x_i e_i$ of X is $\Sigma_1^\infty x_i e_i^{**}$, since for any $x^* = \Sigma_1^\infty \xi_i e_i^*$ in X^* we have

$$\left(\sum_{i=1}^\infty x_i e_i^{**}, \sum_{i=1}^\infty \xi_i e_i^*\right) = \sum_{i=1}^\infty x_i \xi_i = \left(\sum_{i=1}^\infty \xi_i e_i^*, \sum_{i=1}^\infty x_i e_i\right).$$

Therefore $\{e_n^{**}\}$ spans X^{**}. This implies $\{e_n^*\}$ is shrinking, which implies $\{e_n^{**}\}$ is boundedly complete. But $Te_n = e_n^{**}$ for the natural mapping T of X onto X^{**}, so $\{e_n\}$ also is boundedly complete.

Although we already know that c_0 is not reflexive because $(c_0)^{**}$ is l_∞, it is interesting that this also follows from Theorem 6.2 and the natural basis of c_0 not being boundedly complete. The space l_1 is not reflexive, since its natural basis is not shrinking. Actually, each nonreflexive space has a subspace with a basis that is not boundedly complete and a subspace with a basis that is not shrinking [**17**, Theorem 3 (I and II)].

We have seen that the natural basis of l_p is both shrinking and boundedly complete if $1 < p < \infty$. Therefore all l_p spaces with $1 < p < \infty$ are reflexive. In particular, Hilbert space is reflexive.

7. Unconditional Bases. The natural basis of c_0 and the natural basis of any l_p space $(1 < p < \infty)$ are *unconditional*. That is, if $\{e_n\}$ is one of these bases and $x = \Sigma_1^\infty x_i e_i$, then $\Sigma_1^\infty x_{\pi(i)} e_{\pi(i)}$ converges and has sum x if π is any permutation of the positive integers.

In finite-dimensional spaces, a series $\Sigma_1^\infty u_n$ converges unconditionally if and only if $\Sigma_1^\infty \|u_n\| < \infty$. However, in every infinite-dimensional space there is a series that converges unconditionally but not absolutely. In fact, if $\{\alpha_n\}$ are positive numbers with $\Sigma_1^\infty \alpha_n^2 < \infty$, then there is an unconditionally convergent series $\Sigma_1^\infty u_n$ in X such that $\|u_n\| = \alpha_n$ for each n (see [**12**], [**23**(I), Theorem 1.c.2]). For example, in Hilbert space with the orthonormal basis $\{e_n\}$, the series $\Sigma_1^\infty e_n/n$

converges unconditionally and

$$\sum_1^{\infty} \|e_n/n\| = \sum_1^{\infty} 1/n = \infty.$$

One might say that a basis $\{e_n\}$ is *absolutely convergent* if $\Sigma_1^{\infty} x_i e_i$ converges only if $\Sigma_1^{\infty} \|x_i e_i\| < \infty$. The theory of such bases is not very interesting, since any infinite-dimensional space that has such a basis is isomorphic to l_1. In contrast, the theory of unconditional bases is very extensive and interesting. We will discuss some of the most interesting and accessible theory.

The proof of the next theorem is very similar to that of Theorem 3.1 and will not be given. It makes use of a new norm $|||\ \ |||$, defined by letting

$$|||x||| = \sup\left\{\left\|\sum_{i\in A}^{\infty} x_i e_i\right\|\right\} \quad \text{if } x = \sum_{i=1}^{\infty} x_i e_i,$$

where the supremum is over all finite subsets A of the positive integers.

THEOREM 7.1. *If* $\operatorname{lin}\{e_n\}$ *is dense in* X *and no* e_n *is* 0, *then* $\{e_n\}$ *is an unconditional basis for* X *if and only if there is a positive number* K *such that, for all nonempty finite subsets* A *and* B *of the positive integers and scalars* $\{a_n\}$,

$$K\left\|\sum_{i\in A\cup B} a_i e_i\right\| \geqslant \left\|\sum_{i\in A} a_i e_i\right\|. \tag{14}$$

The least number K' that can be used for K in (14) is the *unconditional-basis constant* of $\{e_n\}$. The *unconditional constant* of $\{e_n\}$ is the least number K'' such that

$$K''\left\|\sum_{i\in A\cup B} a_i e_i\right\| \geqslant \left\|\sum_{i\in A} a_i e_i - \sum_{i\in B} a_i e_i\right\|, \tag{15}$$

for all nonempty disjoint finite subsets A and B of the positive integers and all scalars $\{a_n\}$. The triangle inequality can be used to show that $K'' \leqslant 2K'$ and $K' \leqslant \frac{1}{2}(1 + K'')$. The latter implies $K' \leqslant K''$.

There is a simple basis $\{v_n\}$ in c_0 that is not unconditional, for which $v_n = (1,\ldots, 1, 0, 0,\ldots)$ has 1 in each of the first n positions and 0 thereafter. It is an easy exercise to show that the basis constant is 2 and that

$$x = \sum_1^{\infty} (a_n - a_{n+1}) v_n \text{ if } x = (a_1, a_2, a_3, \ldots).$$

This basis is not unconditional, since

$$\left\|\sum_{k=1}^{2n} (-1)^{k+1} v_k\right\| = 1 \text{ and } \left\|\sum_{j=1}^{n} (-1)^{2j} v_{2j-1}\right\| = \left\|\sum_{j=1}^{n} v_{2j-1}\right\| = n$$

for all n, so (14) is not satisfied for any K.

It will be very useful to know that, if $|\theta_i| \leqslant 1$ for each i and each θ_i is real, then K'' also satisfies

$$K''\left\|\sum_{i=1}^{\infty} a_i e_i\right\| \geqslant \left\|\sum_{i=1}^{\infty} \theta_i a_i e_i\right\|. \tag{16}$$

If the scalars $\{\theta_i\}$ are complex, then K'' must be replaced by $2K''$. To verify (16), let x^* be a linear functional with $\|x^*\| = 1$ and

$$\left(x^*, \sum_{i=1}^{\infty} \theta_i a_i e_i\right) = \sum_{i=1}^{\infty} \theta_i a_i (x^*, e_i) = \left\|\sum_{i=1}^{\infty} \theta_i a_i e_i\right\|. \tag{17}$$

If A is the set of all i for which $a_i(x^*, e_i) > 0$ and B is the set of all i for which $a_i(x^*, e_i) < 0$,

then it follows from (17) that

$$\left\|\sum_{i=1}^{\infty}\theta_i a_i e_i\right\| \leqslant \sum_{i=1}^{\infty}|a_i(x^*,e_i)| \leqslant \sum_{i\in A} a_i(x^*,e_i) - \sum_{i\in B} a_i(x^*,e_i)$$
$$= \left(x^*, \sum_{i\in A} a_i e_i - \sum_{i\in B} a_i e_i\right) \leqslant \left\|\sum_{i\in A} a_i e_i - \sum_{i\in B} a_i e_i\right\|$$
$$\leqslant K''\left\|\sum_{i=1}^{\infty} a_i e_i\right\|.$$

The complex case follows from (16) with an application of the triangle inequality.

There are many characterizations of unconditional bases. The next theorem gives some of these.

THEOREM 7.2. *Each of the following is a necessary and sufficient condition for a basis $\{e_n\}$ of a Banach space X to be unconditional.*

(i) *There is a number C such that $C\|\Sigma_1^\infty a_i e_i\| \geqslant \|\Sigma_1^\infty b_i e_i\|$ if $|a_i| \geqslant |b_i|$ for each i.*

(ii) *If $|a_i| \geqslant |b_i|$ for each i, then the convergence of $\Sigma_1^\infty a_i e_i$ implies the convergence of $\Sigma_1^\infty b_i e_i$.*

(iii) *If A is a subset of the positive integers and $\Sigma_1^\infty a_i e_i$ converges, then $\Sigma_{i\in A} a_i e_i$ converges.*

(iv) *For each permutation π of the positive integers, the sequence $\{e_{\pi(i)}\}$ is a basis for X.*

We know that an unconditional basis satisfies (i), since (16) implies C can be K'' for real scalars or $2K''$ for complex scalars. Condition (ii) can be proved by using (i) to show that the Cauchy-convergence property of $\Sigma_1^\infty a_i e_i$ implies the Cauchy-convergence property for $\Sigma_1^\infty b_i e_i$. Condition (iii) is the special case of (ii) for which each b_i either is a_i or 0.

Intuitively, it seems that (iv) is formally almost the same as the definition of unconditionality. Thus it should not be surprising that some argument is needed to show that (iii) implies (iv). We suppose there is a permutation π for which $\{e_{\pi(n)}\}$ is not a basis. Then there is no n for which $\{e_{\pi(i)}: i \geqslant n\}$ is a basis for the closure of its linear span, so no such sequence has a basis constant. Therefore, we can by induction find an increasing sequence of integers $\{p_n\}$ associated with sequences $\{u_n\}$ and $\{v_n\}$ for which $\|u_n\| = 1$ and each u_n belongs to

$$\mathrm{lin}\{e_i: p_{2n-1} < i \leqslant p_{2n}\},$$

each v_n belongs to

$$\mathrm{lin}\{e_i: p_{2n} < i \leqslant p_{2n+1}\},$$

and $\|u_n + v_n\| < 2^{-n}$. Then $\Sigma_1^\infty(u_n + v_n)$ converges and therefore is the expansion of some member of X in the basis $\{e_n\}$. However, $\Sigma_1^\infty u_n$ is not convergent, which contradicts (iii).

The only link still missing in the proof of Theorem 7.2 is to show that $\{e_n\}$ is unconditional if (iv) is satisfied. For this, we observe that if $x \in X$, then for each permutation π there is a representation of x as $\Sigma_1^\infty a_{\pi(i)} e_{\pi(i)}$. Since $(e_{\pi(n)}^*, x) = a_{\pi(n)}$ for each n, the coefficient of each e_n is determined only by x and is independent of the permutation, so $\{e_n\}$ is unconditional.

Because of Theorem 7.2 (i), a space X with an unconditional basis $\{e_n\}$ can be given an equivalent norm $|||\ \ |||$ for which $|||\ \Sigma_{i=1}^\infty x_i e_i|||$ is the supremum of $|||\ \Sigma_{i=1}^\infty a_i e_i|||$ for all $\{a_n\}$ for which $|a_i| \leqslant |x_i|$ for each i. This norm has the property that

$$|||\sum_{i=1}^{\infty} a_i e_i||| \geqslant |||\sum_{i=1}^{\infty} b_i e_i||| \quad \text{if} \quad |a_i| \geqslant |b_i| \text{ for each } i. \tag{18}$$

With this norm, X becomes a *commutative Banach algebra* if we define the product xy to be $\Sigma_1^\infty x_i y_i e_i$ when $x = \Sigma_1^\infty x_i e_i$ and $y = \Sigma_1^\infty y_i e_i$, since then $\|xy\| \leqslant \|x\|\ \|y\|$. Also, X becomes a *Banach lattice* (or *complete normed vector lattice*) if $x \leqslant y$ means $x_i \leqslant y_i$ for each i, and $x \vee y$ and $x \wedge y$ are $\Sigma_1^\infty(x_i \vee y_i)e_i$ and $\Sigma_1^\infty(x_i \wedge y_i)e_i$, respectively. Then

$$|x| = \sum_{1}^{\infty} |x_i| e_i \text{ and } \|x\| \leqslant \|y\| \text{ if } |x| \leqslant |y|.$$

8. Unconditional bases and reflexivity. For a subspace X of a space with an unconditional basis, the nonreflexive spaces c_0 and l_1 illustrate in a very real sense the only ways in which X can fail to be reflexive.

THEOREM 8.1. *If X is a subspace of a space with an unconditional basis, then X is reflexive unless it contains a subspace isomorphic to c_0 or a subspace isomorphic to l_1.*

The proof of this theorem was given first for the case X itself has an unconditional basis [**15**, p. 521]. Because of Theorem 6.2, this proof needed only Lemmas 8.3 and 8.4, below. This was extended to Theorem 8.1 in [**4**]. Before stating and proving these lemmas, let us discuss some consequences of Theorem 8.1.

COROLLARY 8.2. *If X is a subspace of a space with an unconditional basis, then X is reflexive if X^{**} is separable.*

This is an easy corollary of Theorem 8.1, since with care one can use the fact that l_∞ is not separable to show that if X contains a subspace isomorphic to either c_0 or l_1, then X^{**} is not separable.

The space J we have discussed is isometric to J^{**}, so J^{**} also is separable. Since J is not reflexive, J not only does not have an unconditional basis but J is not isomorphic to any subspace of a space with an unconditional basis. Every separable Banach space can be embedded isometrically in $C[0, 1]$ [2, p. 187], so $C[0, 1]$ has no unconditional basis (also see [**21**]).

The space $L_1[0, 1]$ has no unconditional basis, but this is more difficult to show and involves methods and concepts we do not wish to describe in detail. For example, one can show that $L_1[0, 1]$ contains no subspace isomorphic to c_0. Then the next lemma implies any unconditional basis is boundedly complete, and therefore that $L_1[0, 1]$ is isomorphic to a dual if it has an unconditional basis. There are several reasons why this is impossible; *e.g.*, that $L_1[0, 1]$ fails to have the Radon-Nikodým property and all separable duals have the Radon-Nikodým property. A short, but not easy, proof that $L_1[0, 1]$ is not isomorphic to a subspace of any space with an unconditional basis is given in [**23**(I), p. 24].

The trigonometric system is an unconditional basis for $L_p[0, 1]$ if and only if $p = 2$ [**29**(I), p. 428]. However, the Haar system we have discussed is an unconditional basis if $1 < p < \infty$ [29(I), p. 407]. Therefore $L_p[0, 1]$ is reflexive if $1 < p < \infty$, since it contains no subspaces isomorphic to c_0 or l_1.

LEMMA 8.3. *If a Banach space X has an unconditional basis $\{e_n\}$ that is not boundedly complete, then X has a subspace isomorphic with c_0.*

To help in understanding this lemma, there is no loss of generality if we assume that the norm of X satisfies (18). If $\{e_n\}$ is not boundedly complete, then there is a nonconvergent series $\sum_1^\infty a_i e_i$ and a number M such that $\|\sum_1^n a_i e_i\| \leqslant M$ for all n. Since the series is not convergent, there is a positive number Δ and an increasing sequence of integers $\{p_n\}$ for which $\|w_n\| > \Delta$ if $w_n = \sum_{p_n}^{p_{n+1}-1} a_i e_i$. For any scalars $\{c_k\colon 1 \leqslant k \leqslant n\}$, we can use (18) to see that

$$\|c_j w_j\| \leqslant \left\| \sum_{k=1}^{n} c_k w_k \right\| \leqslant \left\| \sum_{k=1}^{n} w_k \right\| \sup\{|c_k|\} \quad \text{for each } j.$$

This implies $\Delta \cdot \sup\{|c_j|\} \leqslant \left\| \sum_1^n c_k w_k \right\| \leqslant M \cdot \sup\{|c_k|\}$, so c_0 is isomorphic with $\mathrm{cl}[\mathrm{lin}\{w_k\}]$.

LEMMA 8.4. *If X has an unconditional basis $\{e_n\}$ that is not shrinking, then X has a subspace isomorphic with l_1.*

Again, let us assume that the norm of X satisfies (18). If $\{e_n\}$ is not shrinking, then there is a linear functional x^* with $\|x^*\| = 1$ and $\lim_{n\to\infty}\|x^*\|_n > 0$. If we denote this limit by θ, then for any positive δ we can choose an increasing sequence of positive integers $\{p_n\}$ and a sequence $\{w_n\}$, which together have the properties: (i) $\|w_n\| = 1$ for each n; (ii) each w_k belongs to $\text{lin}\{e_i : p_k \leqslant i < p_{k+1}\}$; (iii) $(x^*, w_k) > \theta - \delta$ for each k. Then for any scalars $\{c_k\}$,

$$(\theta - \delta)\sum_{k=1}^{n}|c_k| \leqslant \left(x^*, \sum_{k=1}^{n}|c_k|w_k\right) \leqslant \left\|\sum_{k=1}^{n}|c_k|w_k\right\|$$
$$= \left\|\sum_{k=1}^{n}c_k w_k\right\| \leqslant \sum_{k=1}^{n}|c_k|,$$

so l_1 is isomorphic with cl[lin$\{w_k\}$].

9. Existence of bases. There are two classical spaces that were not known to have bases for many years: The space $C^k(I^n)$ of k-times differential functions of n variables with values in $[0, 1]$ and an appropriate norm (bases are given in [**8**] and [**28**]), and the disc algebra of functions analytic on $|z| < 1$ and continuous on $|z| \leqslant 1$ with $\|f\|$ the maximum of $|f(z)|$ (a basis is given in [**6**]). It also was an open question for many years whether every separable Banach space has a basis [**2**, p. 111].

The first proof that there are separable spaces without bases was given by Enflo [**13**]. It now is known that every l_p space with $p \neq 2$ has a subsapce without a basis [**30**]. Actually, each of these examples also fails the *approximation property*. A space having the *approximation property* means that, for each positive ε and compact subset K, there is a bounded linear mapping T of the space into itself whose range is finite-dimensional and for which $\|Tx - x\| < \varepsilon$ if $x \in K$. Any space with a basis has the approximation property. To see this, suppose K is an arbitrary compact subset of a space with a basis $\{e_i\}$ and ε is a positive number. For each n, let P_n be the projection defined by

$$P_n\left(\sum_1^\infty x_i e_i\right) = \sum_1^n x_i e_i,$$

and let W_n be the set of all x for which $\|x - P_m x\| < \varepsilon$ if $m \geqslant n$. Since $\{\|P_n\| : n \geqslant 1\}$ is bounded (Theorem 3.1), each W_n is open and it follows from $W_n \subset W_{n+1}$ and $K \subset U_1^\infty W_n$ that there is an N for which $K \subset W_N$. Then $\|P_N x - x\| < \varepsilon$ if $x \in K$.

Before Enflo's example, Grothendieck ([**14**], [**23**(I), p. 35], [**29**(II), p. 718]) had given many conjectures equivalent to the conjecture that every Banach space has the approximation property, each of which is known now to be false; *e.g.*,

(A) $\Sigma_1^\infty a_{nn} = 0$ if the infinite matrix $A = (a_{ij})$ has the properties that $\lim_{j\to\infty} a_{ij} = 0$ for each i, $\Sigma_1^\infty \max\{|a_{ij}| : j \geqslant 1\} < \infty$, and $A^2 = 0$.

(B) $\int_0^1 K(t, t)\,dt = 0$ if K is a continuous function on $[0, 1] \times [0, 1]$ for which $\int_0^1 K(x, t)K(t, y)\,dt = 0$ for all x and y.

(C) For each continuous function f on the square $[0, 1] \times [0, 1]$ and each positive ε, there are numbers $\xi_1, \xi_2, \ldots, \xi_n$ and $\eta_1, \eta_2, \ldots, \eta_n$ in $[0, 1]$ and scalars $c_1, c_2, \ldots, c_n$ such that, for all (x, y),

$$\left|f(x, y) - \sum_{i=1}^{n} c_i f(x, \eta_i) f(\xi_i, y)\right| < \varepsilon.$$

Contradiction of (A) occurs explicitly in some proofs that there are subspaces of c_0 and l_p $(2 < p < \infty)$ that fail the approximation property ([**9**], [**23**(I), p. 87]). Developments since Enflo's example are summarized in [**29**(II), pp. 720-721].

Since there are separable spaces without bases, it is natural for mathematicians to be curious about the strongest similar structure that might be found for every separable space. It has been shown [**25**] that for every separable X and any positive ε there is a sequence $\{(e_n, e_n^*); i \geqslant 1\}$ for which:

(i) $\|e_n\| = 1$ and $\|e_n^*\| < 1 + \varepsilon$ for each n;
(ii) $(e_i^*, e_j) = 1$ if $i = j$ and 0 otherwise;
(iii) $\mathrm{cl}[\mathrm{lin}\{e_n\}] = X$;
(iv) $\{e_n^*\}$ is *total*; *i.e.*, $x = 0$ if $(e_n^*, x) = 0$ for each n.

It is not known if this is true for $\varepsilon = 0$, except when the space is finite-dimensional [**23**(I), p. 16, 1.c.3].

Each infinite-dimensional subspace of Hilbert space has an orthonormal basis. Even after Enflo's example was known, it still was an open question whether a separable space is isomorphic to Hilbert space if each subspace has a basis. Now this is known to be false [**19**, Example 2.2]. However, it is true that if X is infinite-dimensional, then X has an infinite-dimensional subspace with a basis. This is easy to show. For any $K > 1$, we choose a sequence $\{(e_n, S_n)\}$ inductively so that each S_n is a finite subset of the unit ball of X^*, each $S_n \subset S_{n+1}$, and:

(a) $\|x\| \leqslant K \cdot \sup\{|(x^*, x)|: x^* \in S_n\}$ if $x \in \mathrm{lin}\{e_i: i \leqslant n\}$;
(b) $(x^*, e_k) = 0$ if $x^* \in S_n$ and $k > n$.

To see that this is possible, we let e_1 be any nonzero member of X and let S_1 consist of one x^* for which $\|x^*\| = 1$ and $(x^*, e_1) = \|e_1\|$. If $\{(e_k, S_k): k < n\}$ has been determined, we choose a nonzero e_n in the intersection of the null spaces of the members of S_{n-1} and then choose S_n to contain S_{n-1} and suitable other members of X^* so that (a) is satisfied. Now we observe that, for all scalars $\{a_n\}$ and positive integers n and p,

$$K\left\|\sum_{i=1}^{n+p} a_i e_i\right\| \geqslant \left\|\sum_{i=1}^{n} a_i e_i\right\|,$$

since there is a member x^* of S_n for which

$$\left\|\sum_1^n a_i e_i\right\| \leqslant K \cdot \left|\left(x^*, \sum_1^n a_i e_i\right)\right|$$

and

$$K\left\|\sum_{i=1}^{n+p} a_i e_i\right\| \geqslant K \cdot \left|\left(x^*, \sum_1^{n+p} a_i e_i\right)\right| = K \cdot \left|\left(x^*, \sum_1^n a_i e_i\right)\right| \geqslant \left\|\sum_1^n a_i e_i\right\|.$$

Therefore $\{e_n\}$ is a basic sequence with basis constant K.

The situation for unconditional bases is much more complicated. If X is an infinite-dimensional subspace of a space with an unconditional basis, then it is very easy to show that some infinite-dimensional subspace of X has an unconditional basis. However, one of the most important unsolved conjectures in Banach space theory is whether *each infinite-dimensional space contains an infinite-dimensional subspace with an unconditional basis*.

It follows from Theorem 8.1 that the truth of this conjecture would imply the truth of the conjecture that *each infinite-dimensional space has an infinite dimensional subspace that either is reflexive or isomorphic to one of c_0 or l_1*.

Since X is isometric with X^{**} if X is reflexive, the truth of this conjecture would imply the truth of the conjecture that *each infinite-dimensional space has an infinite-dimensional subspace that either is isomorphic to c_0 or l_1, or has a separable second dual*. Actually, this conjecture is equivalent to the preceding one, since any infinite-dimensional space X with X^{**} separable contains a reflexive subspace [**23**(I), p. 14].

Since the dual of c_0 is l_1 and any space with a separable second dual has a separable first dual, the truth of the last conjecture would imply the truth of the conjecture that *each infinite-dimensional space has an infinite-dimensional subspace that either is isomorphic with l_1 or has a separable dual*. It is not known whether this conjecture is true.

References

1. A. Andrew, James' quasi-reflexive space is not isomorphic to any subspace of its dual, Israel J. Math., 38 (1981) 276–282.

2. S. Banach, Théorie des Opérations Linéaires, Warszawa, 1932. Reprinted by Chelsea, New York, 1955.

3. S. Bellenot, Transfinite duals of quasi-reflexive Banach spaces, to appear in Trans. Amer. Math. Soc.

4. C. Bessaga, and A. Pełczyński, On subspaces of a space with an absolute basis, Bull. Acad. Sci. Pol., 6 (1958) 313–315.

5. ______, Banach spaces non-isomorphic to their Cartesian squares, Bull. Acad. Sci. Pol., 8 (1960) 77–80.

6. S. V. Botschkariev, Existence of a basis in the space of analytic functions, and some properties of the Franklin system, Mat. Sb., 24 (1974) 1–16.

7. P. G. Casazza, James' quasi-reflexive space is primary, Israel J. Math., 26 (1977) 294–305.

8. Z. Ciesielski and J. Domsta, Construction of an orthonormal basis in $C^m(I^d)$ and $W_p^m(I^d)$, Studia Math., 41 (1972) 211–224.

9. A. M. Davie, The approximation problem for Banach spaces, Bull. London Math. Soc., 5 (1973) 261–266.

10. M. M. Day, Normed Linear Spaces, 3rd ed., Springer, Berlin-Heidelberg-New York, 1973.

11. J. Dieudonné, Complex structures on real Banach spaces, Proc. Amer. Math. Soc., 3 (1952) 162–164.

12. A. Dvoretzky and C. A. Rogers, Absolute and unconditional convergence in normed linear spaces, Proc. Nat. Acad. Sci. U.S.A., 36 (1950) 192–197.

13. Per Enflo, A counterexample to the approximation property, Acta Math., 130 (1973) 309–317.

14. A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, Mem. Amer. Math. Soc., 16 (1955).

15. R. C. James, Bases and reflexivity of Banach spaces, Ann. of Math., 52 (1950) 518–527.

16. ______, A non-reflexive Banach space isometric with its second conjugate space, Proc. Nat. Acad. Sci. U.S.A., 37 (1951) 174–177.

17. ______, Super-reflexive Banach spaces, Canad. J. Math., 24 (1972) 896–904.

18. ______, Banach spaces quasi-reflexive of order one, Studia Math., 60 (1977) 157–177.

19. W. B. Johnson, Banach spaces all of whose subspaces have the approximation property, Special Topics of Applied Mathematics (Proc. Sem. Bonn 1979), North-Holland, Amsterdam, 1980.

20. W. B. Johnson, H. P. Rosenthal, and M. Zippin, On bases, finite-dimensional decompositions and weaker structures in Banach spaces, Israel J. Math., 9 (1971) 488–506.

21. S. Karlin, Bases in Banach spaces, Duke Math. J., 15 (1948) 971–985.

22. Bor-Luh Lin and R. H. Lohman, On generalized James quasi-reflexive Banach spaces, Bull. Inst. Math. Acad. Sinica, 8 (1980) 389–399.

23. J. Lindenstrauss and L. Tzafriri, Classical Banach Spaces, vols. 1 and 2, Springer, Berlin-Heidelberg-New York, 1977 and 1979.

24. J. T. Marti, Introduction to the Theory of Bases, Springer, Berlin-Heidelberg-New York, 1969.

25. A. Pełczyński, All separable Banach spaces admit for each $\varepsilon > 0$ fundamental and total biorthogonal sequences bounded by $1 + \varepsilon$, Studia Math., 55 (1976) 295–304.

26. J. Schauder, Zur theorie stetiger Abbildungen in Funktionalräumen, Math. Z., 26 (1927) 47–65.

27. ______, Eine Eigenschaft des Haarschen Orthogonalsystems, Math. Z., 28 (1928) 317–320.

28. S. Schonefeld, Schauder bases in the Banach spaces $C^k(T^q)$, Trans. Amer. Math. Soc., 165 (1972) 309–318.

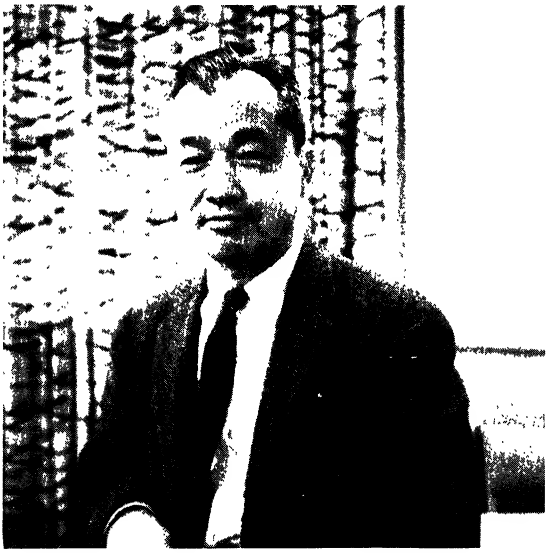
29. I. Singer, Bases in Banach Spaces, vols. 1 and 2, Springer, Berlin-Heidelberg-New York, 1970 and 1981.

30. A. Szankowski, Subspaces without the approximation property, Israel J. Math. 30 (1978) 123–129.

MISCELLANEA

85. We ought to speak the truth; we ought not by silence to countenance a lie; but that does not mean that we must say all we know every time we speak. In our mathematical teaching we often try to do so to the confusion of our pupils.... Greater generality sometimes leads to simplification, but sometimes it does not. It calls for the nicest judgment to decide when to generalize a theorem and when to be content for the time being with a more restricted form.

—G. B. Jeffery, Presidential address (1937) to the London Mathematical Society: *J. London Math. Soc.*, 13 (1938) 72.



Perhaps the ergodic theorist and probabilist is too well known to cause any recognition difficulty, but how about the other one, probably the best known mathematician born in Johannesburg? The answers are on p. 653.

$$1 + l - w = 1 + e(2\pi/3) - e(\pi/3) = 0.$$

Therefore, the identity above holds.

Now equations (4.12), (4.16), (4.17), (4.19) and (4.21) combined are equivalent to

$$a_1 = a_3 = b_1 = b_3 = c_1 = c_3 = 1/3, \quad w = e(\pi/3). \tag{4.22}$$

Since these values are permissible values of the parameters (cf. (2.10) and (2.11)), we can conclude from the discussions above that for equation (2.13) to become an identity in β, γ, it is necessary that (4.22) holds. And this means, in view of what was said at the end of §2, that the "only if" part of Theorem C is proved.

The "if" part of Theorem C is essentially Morley's theorem itself for which many proofs are known. However, we would get not only a new proof of Morley's theorem but also a check on our computation by verifying that equation (2.13) actually becomes an identity in β, γ when the parameters a_1, a_3, b_1, b_3, c_1, c_3 and w have the values given in (4.22). It turns out that we then have

$$\begin{aligned} &\theta_1 = \theta_2 = \theta_3 = \theta_4 = -\tfrac{1}{3}, \\ &\theta_5 = \theta_6 = \tfrac{1}{3}, \\ &\theta_7 = \theta_8 = \theta_9 = \theta_{10} = \theta_{11} = 0, \\ &\theta_{12} = \theta_{13} = \theta_{14} = \theta_{15} = -1, \\ &\theta_{16} = \theta_{17} = \theta_{18} = -\tfrac{2}{3}, \end{aligned} \tag{4.23}$$

and

$$w = e(\pi/3), \quad k = e(-2\pi/3), \quad l = e(2\pi/3);$$

and consequently, equation (2.13) may be rewritten as

$$\tilde{u}_1(\gamma)e(-\beta/3) + \tilde{u}_2(\gamma)e(\beta/3) + \tilde{u}_3(\gamma)e(0) + \tilde{u}_4(\gamma)e(-\beta) + \tilde{u}_5(\gamma)e(-2\beta/3) = 0. \tag{4.24}$$

Now, we can easily verify that the $\tilde{u}_\lambda(\gamma)$ in (4.24) are each a linear combination of $e(2\gamma/3)$, $e(\gamma)$, $e(\gamma/3)$, $e(0)$, $e(-\gamma/3)$ whose coefficients, on account of (4.23), are all equal to zero. Therefore, $\tilde{u}_\lambda(\gamma)$ are all identically zero in γ, and consequently, equation (2.13) $\equiv$ (4.24) is identically zero in β, γ. Theorem C is thus completely proved.

References

1. D. J. Kleven, Morley's theorem and a converse, this MONTHLY, 85 (1978) 100–105.
2. C. O. Oakley and J. C. Baker, The Morley trisector theorem, this MONTHLY, 85 (1978) 737–745.
3. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.
4. Y. C. Wong, Some properties of the triangle, this MONTHLY, 48 (1941) 530–535.
5. Y. C. Wong, Some extensions of the Douglas-Neumann theorem for concentric polygons, this MONTHLY, 75 (1968) 470–482.

ANSWERS TO "PHOTOS" ON PAGE 641

Above: Shizuo Kakutani; below: Claude Chevalley. Both pictures were taken in 1965.

A STRONG CONVERSE OF MORLEY'S TRISECTOR THEOREM

YUNG-CHOW WONG AND KAI-MAN TSANG
Department of Mathematics, University of Hong Kong, Hong Kong

Dedicated to Professor Shiing-Shen Chern for his outstanding work and his invaluable service to the community of mathematicians.

1. Introduction. The celebrated trisector theorem in elementary geometry discovered by F. Morley about 1889 can be stated as follows.

THEOREM A. *The three points of intersection of the adjacent trisectors of the angles of any triangle form an equilateral triangle.*

Many papers have been written about this theorem, but it appears from the very comprehensive list of references contained in Oakley and Baker's paper [**2**] that the question of its converse was not considered until quite recently. In 1978, D. J. Kleven [**1**] proved the following theorem which we restate as follows:

THEOREM B. *Let a_1, b_1, c_1 be given real numbers lying strictly between* 0 *and* 1/2. *Take an arbitrary nondegenerate triangle ΔABC, and from its vertex A, draw two straight lines inside ΔABC so that they form with the sides AB and AC angles equal to a_1 times the angle $\angle A$, etc. Denote by P the point of intersection of the two straight lines adjacent to the side BC of ΔABC, etc. Then ΔPQR will be equilateral for all nondegenerate ΔABC iff $a_1 = b_1 = c_1 = 1/3$.*

Here and in Theorem C below, "etc." means cyclic permutations of the letters A, B, C; a_1, b_1, c_1; P, Q, R.

Kleven called his theorem a converse to Morley's theorem and said that it "shows that we cannot get another theorem like Morley's by dividing the angles in some other way," (see [**1**], p. 101, lines 7 and 6). Unfortunately, his assertion: "we cannot... way," though it turns out to be true, does not follow from his theorem. This is because his theorem specifies from the outset that the angles of the arbitrary starting triangle are each to be so divided into three parts that the two nonadjacent parts of the angle are equal. Thus, strictly speaking, Kleven's theorem is only a weak converse of Morley's theorem, as a true converse should not contain such specifications.

In this paper, we shall settle the question of the converse to Morley's theorem by proving a theorem in which no such specifications are made and, moreover, at the outset, we require of the ΔPQR not that it be equilateral, but only that it be similar to a given triangle. Because of this relaxation of the condition on ΔPQR, our theorem may be said to be a strong converse of Morley's theorem.

THEOREM C. *Let a_i, b_i, c_i $(i = 1,2,3)$ be given real numbers lying strictly between* 0 *and* 1 *such that*

$$\Sigma_i a_i = 1, \quad \Sigma_i b_i = 1, \quad \Sigma_i c_i = 1,$$

Yung-Chow Wong graduated from Sun Yat-Sen University in China and received his Ph.D. and D.Sc. degrees at the University of London. After spending several years in the United States doing research and some teaching, he returned to China and taught briefly at his Alma Mater. He then settled down in Hong Kong where he occupied the Chair of Mathematics at the University of Hong Kong for 28 years and was Head of the Department of Mathematics most of that time, also serving on two separate occasions as Dean of the Faculty of Engineering and Pro-Vice-Chancellor. He is now retired. His research interest is in geometry. He has published many papers on differential geometry and a book on linear geometry in Euclidean 4-space. He is also the Founding President of the Southeast Asian Mathematical Society and Honorary President of the recently formed Hong Kong Mathematical Society.

Kai-Man Tsang graduated from the University of Hong Kong and was teaching assistant in the Department of Mathematics for two years. He is now a Ph.D. student at Princeton University. He has published three papers on analytic number theory.

and let ΔLMN be any given triangle. Take an arbitrary nondegenerate ΔABC, and from each of its vertices draw two straight lines inside ΔABC to divide the angles $\angle A$, $\angle B$, $\angle C$ into three parts proportional to a_1, a_2, a_3; b_1, b_2, b_3; c_1, c_2, c_3, respectively. Denote by P the point of intersection of the two straight lines adjacent to the side BC of ΔABC, etc. Then, for all nondegenerate ΔABC with the same orientation, ΔPQR will be similar to ΔLMN iff $a_i, b_i, c_i (i = 1, 2, 3)$ are all equal to $1/3$ and ΔLMN is equilateral and has the same orientation as ΔABC.

Here two nondegenerate triangles are said to have the same orientation if their boundaries are described in the same sense of rotation. Two triangles are said to be similar if they are both nondegenerate and have the same orientation and their corresponding sides are proportional (and this implies that their corresponding angles are equal), or if they are both degenerate and their corresponding sides are proportional.

Unlike Kleven's proof of his theorem, the proof of our theorem will be carried out in the complex plane, using the method developed by Wong in his two earlier papers [**4**], [**5**]. In §2, we express the condition for ΔPQR to be similar to ΔLMN as an equation (see (2.13)) containing the real numbers $a_1, a_3, b_1, b_3, c_1, c_3$ which appear in the statement of Theorem C, a complex number w which determines the shape and orientation of the given ΔLMN, and two real variables β and γ which determine the angles of the arbitrary ΔABC. In §3, we prove some lemmas that will be needed, and finally in §4, we prove Theorem C by finding the values of the parameters a_1, a_3, b_1, b_3, c_1, c_3 and w so that equation (2.13) will become an identity in β, γ.

REMARK 1. The assumption that the real numbers a_i, b_i, c_i all lie strictly between 0 and 1 means that (a) the two straight lines drawn from any vertex of ΔABC do not coincide and (b) none of the straight lines drawn from the vertices coincides with a side of ΔABC. A consequence of (a) is that the points P, Q, R are all distinct, and therefore, we may assume that the vertices of the given ΔLMN are all distinct.

REMARK 2. When we began working on this problem, we purposely relaxed the condition on ΔPQR, hoping to find in the end that ΔPQR might be of a more general shape (which may possibly depend on the values of the parameters a_i, b_i, c_i) than being equilateral, so that we would have not only a proof of the converse to Morley's theorem, but also a generalization. As it has turned out, however, ΔPQR must be equilateral.

2. The Main Equation. From now on we shall assume that everything takes place in the complex plane provided with complex coordinate $z = x + iy$, where $i = \sqrt{-1}$ and x, y are rectangular coordinates. In this section we shall derive an equation expressing analytically the condition for the ΔPQR and ΔLMN in Theorem C to be similar.

The complex coordinate of a point on the unit circle is of the form $\cos\theta + i\sin\theta$, where θ is a real number. For convenience, we shall denote $\cos\theta + i\sin\theta$ by $e(\theta)$ and often call the point with coordinate $e(\theta)$ simply the point $e(\theta)$.

Let $e(\theta)$ and $e(\theta')$ be any two points on the unit circle. Then direct verification will show that the straight line passing through these two points has the equation

$$z + \bar{z}e(\theta + \theta') = e(\theta) + e(\theta'), \tag{2.1}$$

where, as usual, $\bar{z}$ is the conjugate of z. Using (2.1), we can easily prove that the straight line passing through the points $e(\theta)$, $e(\theta')$ and the line passing through the points $e(\phi)$, $e(\phi')$ intersect at the point G whose coordinate g is the conjugate of

$$\bar{g} = \frac{e(\theta) + e(\theta') - e(\phi) - e(\phi')}{e(\theta + \theta') - e(\phi + \phi')}. \tag{2.2}$$

In proving Theorem C, we may assume without loss of generality that the starting nondegenerate ΔABC is positively oriented and its vertices are the points

$$A: e(0) = 1, \quad B: e(\gamma), \quad C: e(\gamma + \alpha) \tag{2.3}$$

on the unit circle, so that $0 < \gamma < \gamma + \alpha < 2\pi$ and the angles $\angle A$, $\angle B$, $\angle C$ of ΔABC are equal to $\alpha/2$, $\beta/2 = (2\pi - \alpha - \gamma)/2$, and $\gamma/2$, respectively. Now draw from each of the vertices A, B, C two straight lines lying inside ΔABC and dividing the angles $\angle A$, $\angle B$, $\angle C$ into three parts which are proportional to a_1, a_2, a_3; b_1, b_2, b_3; c_1, c_2, c_3, respectively. Then as shown in the figure below, these six lines meet the unit circle again at the points

$$\begin{aligned} &A_1 : e(\gamma + \alpha_1), && A_2 : e(\gamma + \alpha_1 + \alpha_2), \\ &B_1 : e(\gamma + \alpha + \beta_1), && B_2 : e(\gamma + \alpha + \beta_1 + \beta_2), \\ &C_1 : e(\gamma_1), && C_2 : e(\gamma_1 + \gamma_2), \end{aligned} \tag{2.4}$$

where

$$\alpha_i = a_i \alpha, \quad \beta_i = b_i \beta, \quad \gamma_i = c_i \gamma \quad (i = 1,2,3). \tag{2.5}$$

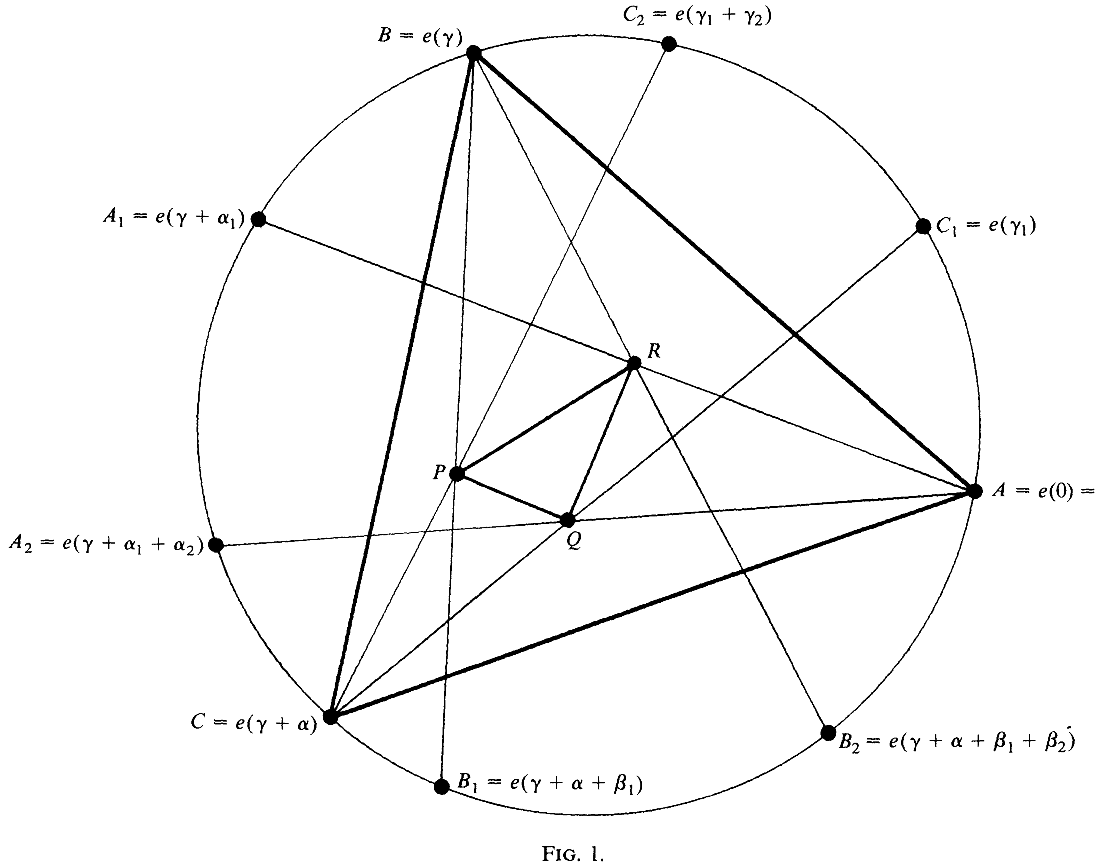


FIG. 1.

Since P is the point at which the straight line passing through the points B, B_1 intersects the straight line passing through the points C, C_2, we can obtain the coordinate p of P by using (2.3), (2.4) and (2.2). Thus putting

$$\theta = \gamma, \quad \theta' = \gamma + \alpha + \beta_1, \quad \phi = \gamma + \alpha, \quad \phi' = \gamma_1 + \gamma_2$$

in (2.2) and using

$$\gamma = \gamma_1 + \gamma_2 + \gamma_3, \quad \alpha + \beta + \gamma = 2\pi, \quad e(2\pi) = 1, \tag{2.6}$$

we get

$$\bar{p} = \frac{e(\gamma) + e(2\pi - \beta + \beta_1) - e(2\pi - \beta) - e(\gamma - \gamma_3)}{e(\alpha + 2\gamma)[e(\beta_1) - e(-\gamma_3)]}$$
$$= \frac{e(-\alpha - 2\gamma)[e(\gamma) + e(-\beta + \beta_1) - e(-\beta) - e(\gamma - \gamma_3)]}{e(\beta_1) - e(-\gamma_3)}$$
$$= \frac{e(\beta) + e(-\gamma + \beta_1) - e(-\gamma) - e(\beta - \gamma_3)}{e(\beta_1) - e(-\gamma_3)}.$$

Hence

$$p = \frac{e(-\beta) - e(\gamma_3 - \beta) + e(\gamma - \beta_1) - e(\gamma)}{e(-\beta_1) - e(\gamma_3)}. \tag{2.7$_1$}$$

Similarly, we can obtain the following coordinates of the points Q and R:

$$q = \frac{1 - e(\alpha_3) + e(-\beta - \gamma_1) - e(-\beta)}{e(-\gamma_1) - e(\alpha_3)}, \tag{2.7$_2$}$$

$$r = \frac{e(-\alpha_1) - 1 + e(\gamma) - e(\gamma + \beta_3)}{e(-\alpha_1) - e(\beta_3)}. \tag{2.7$_3$}$$

Since the ΔLMN in Theorem C can be replaced by any triangle similar to it and since the vertices L, M, N of ΔLMN are all distinct (see Remark 1 at end of §1), we may assume that they are respectively at the points with coordinates w, 0, 1, where w is a suitable complex number $\neq 0$, 1. (Note that w may be a real number, in which case ΔLMN is degenerate but with distinct vertices.) Then, as is well known, ΔPQR is similar to ΔLMN iff

$$(r - p)/(q - p) = (1 - w)/(0 - w),$$

i.e.,

$$-p + (1 - w)q + wr = 0, \tag{2.8}$$

where p, q, r are the coordinates of P, Q, R as given in (2.7).

To express equation (2.8) explicitly, we first substitute (2.5) in (2.7), and then (2.7) in (2.8). Then after replacing α by $2\pi - \beta - \gamma$ and rearranging the terms, we arrive at the explicit equation (2.13) given below, where for convenience of reference we have denoted each line by a number and where

$$k = e(-2a_1\pi), \quad l = e(2a_3\pi) \quad (\text{so that } k \neq 0, l \neq 0); \tag{2.9}$$

$$\begin{gathered} a_1, a_3, b_1, b_3, c_1, c_3 \text{ are real parameters lying strictly between 0 and 1 such that} \\ a_1 + a_3, \quad b_1 + b_3, \quad c_1 + c_3 \quad \text{are all} < 1; \end{gathered} \tag{2.10}$$

$$w \text{ is a complex parameter} \neq 0, 1; \tag{2.11}$$

and

$$\beta, \gamma \text{ are real variables such that } \beta > 0, \quad \gamma > 0, \quad \beta + \gamma < 2\pi. \tag{2.12}$$

Values of the parameters which satisfy (2.10) and (2.11) will be called permissible values.

[1] $e(-b_1\beta)\{we((1 - c_1)\gamma) - we(-c_1\gamma)\}$

[2] $+e(-a_3\beta)\{lwe((1 + c_3 - a_3)\gamma) - lwe((c_3 - a_3)\gamma)\}$

[3] $+e((-a_3 - b_1 + b_3)\beta)\{-l(1 - w)e((1 - a_3)\gamma) + l(1 - w)e(-a_3\gamma)\}$

[4] $+e((a_1 - a_3 - b_1)\beta)\{kle((1 + a_1 - a_3)\gamma) - kle((a_1 - a_3)\gamma)\}$

[5] $+e(a_1\beta)\{ke((1 + a_1 - c_1)\gamma) - k(1 - w)e((a_1 + c_3)\gamma) - kwe((a_1 + c_3 - c_1)\gamma)\}$

[6] $+e(b_3\beta)\{we((1 - c_1 + c_3)\gamma) - e((1 - c_1)\gamma) + (1 - w)e(c_3\gamma)\}$

$$
\begin{array}{ll}
[7] & +e((a_1 - a_3)\beta)\{-kle((a_1 - a_3 + 1)\gamma) + kle((a_1 - a_3 + c_3)\gamma)\} \\
[8] & +e((a_1 - b_1)\beta)\{-ke((a_1 + 1 - c_1)\gamma) + k(1 - w)e(a_1\gamma) + kwe((a_1 - c_1)\gamma)\} \\
[9] & +e((-b_1 + b_3)\beta)\{(1 - w)e((1 - c_1)\gamma) - (1 - w)\} \\
[10] & +e(0\beta)\{-we((1 + c_3 - c_1)\gamma) + we((c_3 - c_1)\gamma)\} \qquad (2.13) \\
[11] & +e((-a_3 + b_3)\beta)\{-lwe((1 - a_3 + c_3)\gamma) + le((1 - a_3)\gamma) - l(1 - w)e((c_3 - a_3)\gamma)\} \\
[12] & +e((a_1 - b_1 - 1)\beta)\{-k(1 - w)e(a_1\gamma) + k(1 - w)e((a_1 - c_1)\gamma)\} \\
[13] & +e((-b_1 + b_3 - 1)\beta)\{(1 - w) - (1 - w)e(-c_1\gamma)\} \\
[14] & +e((a_1 - a_3 - 1)\beta)\{-kle((a_1 - a_3 + c_3)\gamma) + kle((a_1 - a_3))\gamma\} \\
[15] & +e((-a_3 + b_3 - 1)\beta)\{le((c_3 - a_3)\gamma) - le(-a_3\gamma)\} \\
[16] & +e((-a_3 - b_1)\beta)\{-lwe((1 - a_3)\gamma) + lwe(-a_3\gamma)\} \\
[17] & +e((b_3 - 1)\beta)\{-(1 - w)e(c_3\gamma) - we((c_3 - c_1)\gamma) + e(-c_1\gamma)\} \\
[18] & +e((a_1 - 1)\beta)\{k(1 - w)e((a_1 + c_3)\gamma) + kwe((a_1 + c_3 - c_1)\gamma) - ke((a_1 - c_1)\gamma)\} \\
 & = 0.
\end{array}
$$

We have just shown above that ΔPQR is similar to ΔLMN iff equation (2.13) is satisfied. Now since ΔPQR is constructed from ΔABC whose vertices are given in (2.3), ΔPQR is similar to ΔLMN for all nondegenerate positively oriented ΔABC iff equation (2.13) holds for all the values of β, γ satisfying condition (2.12). Furthermore, it is easy to see that since the vertices L, M, N of ΔLMN are at the points with coordinates $w, 0, 1$, ΔLMN is a positively oriented equilateral triangle iff $w = e(\pi/3)$. Therefore, to prove Theorem C is to prove that (i) for equation (2.13) to become an identity in β, γ, it is necessary that the parameters have the values

$$a_1 = a_3 = b_1 = b_3 = c_1 = c_3 = \tfrac{1}{3}, \quad w = e(\pi/3),$$

and (ii) if the parameters have these values, then equation (2.13) becomes an identity in β, γ. This we shall do in the next two sections. As can be expected, our major task will be in proving (i) which is our strong converse of Morley's theorem, whereas for the proof of (ii), which is a slight extension of Morley's theorem, only straightforward computation is required.

3. Some Lemmas. Equation (2.13) is of the form:

$$\Sigma_j e(\theta_j\beta)u_j(\gamma) \equiv \Sigma_j e(\theta_j\beta)\{\Sigma_h v_{jh} e(\phi_{jh}\gamma)\} = 0, \tag{3.1}$$

where, as comparison with (2.13) will show, $j = 1, \ldots, 18$; $h = 1, 2$ or $1, 2, 3$, depending on j; θ_j are linear functions of a_1, a_3, b_1, b_3; v_{jh} are polynomials of degree $\leqslant 2$ in $k = e(-2a_1\pi)$, $l = e(2a_3\pi)$, and w; ϕ_{jh} are linear functions of a_1, a_3, c_1, c_3; and β, γ are real variables on some finite open intervals. Thus, for example,

$$
\begin{array}{lll}
\theta_5 = a_1, & v_{51} = k, & \phi_{51} = 1 + a_1 - c_1, \\
 & v_{52} = -k(1 - w), & \phi_{52} = a_1 + c_3, \\
 & v_{53} = -kw, & \phi_{53} = a_1 - c_1 + c_3.
\end{array}
$$

Now the functions

$$u_j(\gamma) \equiv \Sigma_h v_{jh} e(\phi_{jh}\gamma) \quad \text{and} \quad F_\gamma(\beta) \equiv \Sigma_j u_j(\gamma) e(\theta_j\beta)$$

which appear in (3.1) are all of the type:

$$F(x) \equiv \Sigma_j t_j e(\psi_j x), \tag{3.2}$$

where $j = 1,\ldots, n$; t_j are complex numbers, ψ_j are real numbers, and x is a real variable on some finite open interval. To prove Theorem C we shall need the following lemmas in which $F(x)$ denotes a function of the type (3.2) and $F(x) \overset{x}{\equiv} 0$ means that $F(x)$ is identically zero.

LEMMA 3.1. *If $F(x) \overset{x}{\equiv} 0$, then at least one of the t_j is zero, or at least two of the ψ_j are equal.*

Proof. Differentiating

$$F(x) \equiv \Sigma_j t_j e(\psi_j x) \overset{x}{\equiv} 0$$

$n - 1$ times with respect to x, we get

$$\Sigma_j t_j \psi_j e(\psi_j x) \overset{x}{\equiv} 0,$$

$$\Sigma_j t_j \psi_j^2 e(\psi_j x) \overset{x}{\equiv} 0,$$

$$\cdots$$

$$\Sigma_j t_j \psi_j^{n-1} e(\psi_j x) \overset{x}{\equiv} 0.$$

Since the $e(\psi_j x)$ are all nowhere zero, it follows from the above that the determinant

$$\begin{vmatrix} t_1 & t_2 & \cdots & t_n \\ t_1\psi_1 & t_2\psi_2 & \cdots & t_n\psi_n \\ \vdots & \vdots & & \vdots \\ t_1\psi_1^{n-1} & t_2\psi_2^{n-1} & \cdots & t_n\psi_n^{n-1} \end{vmatrix} = 0,$$

i.e.,

$$\begin{aligned} t_1 t_2 \cdots t_n(\psi_1 - \psi_2) &\cdots (\psi_1 - \psi_n) \\ &\times(\psi_2 - \psi_3)\cdots(\psi_2 - \psi_n) \\ &\cdots \\ &\times(\psi_{n-1} - \psi_n) = 0, \end{aligned}$$

and this proves our lemma.

An easy consequence of Lemma 3.1 is

LEMMA 3.2. *If the ψ_j in $F(x)$ are all distinct, then $F(x) \overset{x}{\equiv} 0$ iff the t_j are all zero.*

LEMMA 3.3. *Suppose the t_j in $F(x)$ are all nonzero. Then a necessary and sufficient condition that $F(x) \overset{x}{\equiv} 0$ is that* (i) *each ψ_j is equal to at least one other ψ_j, and* (ii) *if*

$$\psi_{j_1} = \cdots = \psi_{j_s} (s \geqslant 2)$$

but not equal to any other ψ_j, then

$$t_{j_1} + \cdots + t_{j_s} = 0.$$

Proof. Let us rearrange the terms in $F(x)$ so that

$$\begin{aligned}\psi_1 &= \psi_2 = \cdots = \psi_{n_1}\\ &> \psi_{n_1+1} = \psi_{n_1+2} = \cdots = \psi_{n_1+n_2}\\ &\cdots\\ &> \psi_{n_1+\cdots+n_{d-1}+1} = \cdots = \psi_{n_1+\cdots+n_d} \equiv \psi_n.\end{aligned}$$

Then $F(x) = \Sigma_\lambda \tilde{t}_\lambda e(\tilde{\psi}_\lambda x)$, where $\lambda = 1,\dots, d$, and

$$\tilde{\psi}_1 = \psi_1, \tilde{\psi}_2 = \psi_{n_1+1},\dots, \tilde{\psi}_d = \psi_{n_1+\cdots+n_{d-1}+1};$$

$$\tilde{t}_1 = t_1 + \cdots + t_{n_1}, \tilde{t}_2 = t_{n_1+1} + \cdots + t_{n_1+n_2},$$

$$\dots, \tilde{t}_d = t_{n_1+\cdots+n_{d-1}+1} + \cdots + t_n.$$

Since the $\tilde{\psi}_\lambda$ are all distinct, it follows from Lemma 3.2 that $F(x) \overset{x}{\equiv} 0$ iff the $\tilde{t}_\lambda$ are all zero. This proves (ii). To prove (i) is to prove that n_λ are all $\geqslant 2$. Assume that $n_1 = 1$, then $t_1 = \tilde{t}_1 = 0$. But this contradicts our hypothesis that all t_j are nonzero. Therefore, $n_1 \geqslant 2$ and our lemma is completely proved.

LEMMA 3.4. *Let x be a real variable on some open interval, and let $t_1,\dots, t_n$, $t'_1,\dots, t'_m$ be nonzero complex numbers and $\psi_1,\dots, \psi_n$, $\psi'_1,\dots, \psi'_m$ real numbers such that*

$$\psi_1 > \cdots > \psi_n, \quad \psi'_1 > \cdots > \psi'_m. \tag{3.3}$$

Then

$$\sum_{j=1}^{n} t_j e(\psi_j x) + \sum_{h=1}^{m} t'_h e(\psi'_h x) \overset{x}{\equiv} 0 \tag{3.4}$$

iff $m = n$ and

$$\psi_1 = \psi'_1,\dots, \psi_n = \psi'_n; \quad t_1 + t'_1 = 0,\dots, t_n + t'_n = 0.$$

Proof. Suppose that (3.4) holds. Then by Lemma 3.3, ψ_1 must be equal to at least one of the other numbers in (3.3). But we see from (3.3) that ψ_1 cannot be equal to any other ψ_j; and, if ψ_1 is equal to some ψ'_h other than ψ'_1, then ψ'_1 would be greater than all the other numbers in (3.3) and this is impossible by Lemma 3.3. Therefore, $\psi_1 = \psi'_1$. Similarly, $\psi_2 = \psi'_2$, etc., and our lemma follows from this and Lemma 3.3.

4. Proof of Theorem C. We are now ready to prove our theorem. To apply the lemmas in the last section to equation (2.13), first write it as

$$\sum_j e(\theta_j\beta) u_j(\gamma) \equiv \sum_j e(\theta_j\beta)\Big\{\sum_h v_{jh} e(\phi_{jh}\gamma)\Big\} = 0, \tag{3.1}$$

and note from (2.13) that on account of (2.9), (2.10), and (2.11), the complex numbers v_{jh} and the real numbers ϕ_{jh} in (3.1) have the following properties:

$$v_{jh} \text{ are all nonzero}, \tag{4.1}$$

$$2 > \phi_{jh} > \phi_{jh'} > -1 \text{ whenever } h < h'. \tag{4.2}$$

Then, by Lemma 3.1, none of the coefficients

$$u_j(\gamma) \equiv \sum_h v_{jh} e(\phi_{jh}\gamma) \tag{4.3}$$

of $e(\theta_j\beta)$ in (3.1) can be identically zero for γ on any finite open interval. In fact, we can prove the following.

LEMMA 4.1. *For any set of permissible values of the parameters* a_1, a_3, c_1, c_3 *and* w, *there exists some open interval of* γ *on which the functions* $u_j(\gamma)$ $(j = 1,\ldots, 18)$ *defined by* (4.3) *are all nowhere zero.*

Proof. Consider the analytic function $f_j(z) \equiv \Sigma_h v_{jh} e(\phi_{jh} z)$ defined on the open disk $D: |z| < 2\pi + \varepsilon, \varepsilon > 0$. Since $f_j(\gamma) \equiv u_j(\gamma)$ for γ in the open interval $(0, 2\pi)$ and since $u_j(\gamma)$ is not identically zero, $f_j(z)$ is not identically zero. Therefore, by a well-known theorem in complex analysis (see, for example, Rudin [3, p. 225]), the zero set of $f_j(z)$ has no limit point in D. It follows from this that the zero set of $u_j(\gamma)$ has no limit point in the closed interval $[0, 2\pi]$, and consequently, by the Bolzano-Weierstrass theorem, $u_j(\gamma)$ can at most vanish at a finite number of points in $(0, 2\pi)$. Since there are only a finite number of the functions $u_j(\gamma)$, there are at most a finite number of points in $(0, 2\pi)$ at which one or another of the functions $u_j(\gamma)$ vanishes. This proves our lemma.

An important consequence of Lemma 4.1 and Lemma 3.3 is

LEMMA 4.2. *Suppose that for some set of permissible values of the parameters* a_1, a_3, b_1, b_3, c_1, c_3 *and* w, *equation* (3.1) $\equiv$ (2.13) *becomes an identity in* β, γ. *Then, for each* $j = j_1$, θ_{j_1} *must be equal to some other* θ_j; *moreover, if* $\theta_{j_1} = \theta_{j_2} = \cdots = \theta_{j_s}$ $(s \geqslant 2)$, *but* $\neq$ *any other* θ_j, *then*

$$u_{j_1}(\gamma) + u_{j_2}(\gamma) + \cdots + u_{j_s}(\gamma) \overset{\gamma}{\equiv} 0$$

on some open interval of γ.

Proof. By Lemma 4.1, there exists some open interval $I \subset (0, 2\pi)$ such that, for every $\gamma_0 \in I$, the $u_j(\gamma_0)$ are all nonzero. Suppose that $\Sigma_j e(\theta_j \beta) u_j(\gamma) \overset{\beta,\gamma}{\equiv} 0$. Then $\Sigma_j e(\theta_j \beta) u_j(\gamma_0) \overset{\beta}{\equiv} 0$ for every $\beta \in (0, 2\pi - \gamma_0)$. Therefore, by Lemma 3.3, the θ_j have the property stated and the $u_j(\gamma)$ are such that

$$u_{j_1}(\gamma_0) + u_{j_2}(\gamma_0) + \cdots + u_{j_s}(\gamma_0) = 0.$$

Since this equation holds for every $\gamma_0 \in I$, our lemma is proved.

Now consider the θ_j in (3.1) $\equiv$ (2.13). It can easily be verified that, because of (2.10), the θ_j have the following properties:

$$\begin{aligned} &\theta_6 \neq \theta_1, \theta_2, \theta_3, \theta_9, \ldots, \theta_{18}, \\ &\theta_5 \neq \theta_4, \theta_7, \theta_8, \\ &\theta_4 \neq \theta_5, \theta_7, \theta_8, \end{aligned} \tag{4.4}$$

and

$$\theta_{12} \neq \theta_1, \ldots, \theta_{11}, \theta_{16}, \theta_{18}. \tag{4.5}$$

For example, if $\theta_{12} = \theta_2$, then

$$a_1 - b_1 - 1 = -a_3, \quad \text{i.e.,} \quad b_1 + (1 - a_1 - a_3) = 0,$$

which however contradicts (2.10). Therefore, $\theta_{12} \neq \theta_2$.

By Lemma 4.2, for equation (2.13) to become an identity in β and γ, θ_6 must be equal to at least one of the other θ_j. In view of (4.4), this can possibly happen in one of the following cases only:

Case 1. $\theta_6 = \theta_4$,

Case 2. $\theta_6 = \theta_7$,

Case 3. $\theta_6 = \theta_8$,

Case 4. $\theta_6 = \theta_7 = \theta_8$,

Case 5. $\theta_6 = \theta_5$,

where, *in each case*, θ_6 *is not equal to any other* θ_j.

We shall now show that of these five cases only Case 5 is possible.

Case 1. $\theta_6 = \theta_4 \neq$ any other θ_j.

In this case, Lemma 4.2 tells us that for equation (2.13) to become an identity in β and γ, we must also have

$$u_6 + u_4 = v_{61}e(\phi_{61}\gamma) + v_{62}e(\phi_{62}\gamma) + v_{63}e(\phi_{63}\gamma) + v_{41}e(\phi_{41}\gamma) + v_{42}e(\phi_{42}\gamma) \overset{\gamma}{\equiv} 0.$$

But, by (4.2) and Lemma 3.4, this is impossible because u_6 and u_4 do not have the same number of terms.

Case 2. $\theta_6 = \theta_7 \neq$ any other θ_j.

This case is also impossible because u_6 and u_7 do not have the same number of terms.

Case 3. $\theta_6 = \theta_8 \neq$ any other θ_j.

In this case, Lemma 4.2 demands that

$$\begin{aligned} u_6 + u_8 = {} & we((1 - c_1 + c_3)\gamma) - e((1 + c_1)\gamma) + (1 - w)e(c_3\gamma) \\ & -ke((1 + a_1 - c_1)\gamma) + k(1 - w)e(a_1\gamma) + kwe((a_1 - c_1)\gamma) \overset{\gamma}{\equiv} 0, \end{aligned}$$

and consequently by Lemma 3.4, we must have

$$1 - c_1 + c_3 = 1 + a_1 - c_1, \quad 1 - c_1 = a_1, \quad c_3 = a_1 - c_1.$$

But these equations are equivalent to

$$a_1 = c_3 = 1, \qquad c_1 = 0,$$

which contradict (2.10). Therefore, this case is impossible.

Case 4. $\theta_6 = \theta_7 = \theta_8 \neq$ any other θ_j.

In this case, Lemma 4.2 demands that

$$\begin{aligned} u_6 + u_7 + u_8 = {} & we(\phi_{61}\gamma) - e(\phi_{62}\gamma) + (1 - w)e(\phi_{63}\gamma) \\ & -kle(\phi_{71}\gamma) + kle(\phi_{72}\gamma) \\ & -ke(\phi_{81}\gamma) + k(1 - w)e(\phi_{82}\gamma) + kwe(\phi_{83}\gamma) \overset{\gamma}{\equiv} 0. \end{aligned} \tag{4.6}$$

We shall now show that this is impossible.

First note that $\theta_6 = \theta_7$ means $b_3 = a_1 - a_3$. Combining this with (2.10), we see that the a's and c's which are contained in the ϕ_{jh} appearing in (4.6) above must satisfy the following conditions:

$$\begin{gathered} a_1, \quad a_3, \quad a_1 - a_3, \quad a_1 + a_3, \quad c_1, \quad c_3, \quad c_1 + c_3 \\ \text{all lie strictly between 0 and 1.} \end{gathered} \tag{4.7}$$

Next, since $k \neq 0$, $l \neq 0$ and $w \neq 0, 1$ (cf. (2.9) and (2.11)), the coefficients of $e(\phi_{jh}\gamma)$ in (4.6) are all nonzero. Consequently, by Lemma 3.3,

$$\text{each of the } \phi_{jh} \text{ in (4.6) must be equal to at least one other } \phi_{jh}. \tag{4.8}$$

Now on account of (4.7), the ϕ_{jh} in (4.6) satisfy the following inequalities:

$$\begin{aligned} \phi_{61} &= 1 - c_1 + c_3 > \phi_{62} = 1 - c_1 > \phi_{63} = c_3, \\ \phi_{71} &= 1 + a_1 - a_3 > \phi_{72} = a_1 - a_3 + c_3 > \phi_{63} = c_3, \\ \phi_{81} &= 1 + a_1 - c_1 > \phi_{82} = a_1 > \phi_{83} = a_1 - c_1. \end{aligned} \tag{4.9}$$

It follows from this that

$$\phi_{61}, \phi_{62}, \phi_{71}, \phi_{72} \text{ are all} > \phi_{63},$$

and consequently, ϕ_{63} must be equal to one of $\phi_{81}, \phi_{82}, \phi_{83}$. But if $\phi_{63} = \phi_{81}$ or ϕ_{82}, then ϕ_{83} would be $<$ all the other numbers in (4.9), and this contradicts (4.8). Therefore, $\phi_{63} = \phi_{83}$ which is then $<$ all the other numbers in (4.9), and in this case,

$$a_1 = c_1 + c_3. \tag{4.10}$$

Combining (4.10) with (4.9) we see that

$$\phi_{81} = 1 + a_1 - c_1 = 1 + c_3 > \phi_{61},$$

and consequently,

$$\phi_{81} > \text{all the } \phi_{61}, \phi_{62}, \phi_{63}, \phi_{82}, \phi_{83}.$$

Therefore, ϕ_{81} must be equal to ϕ_{71} or ϕ_{72}. But if $\phi_{81} = \phi_{72}$, then ϕ_{71} would be $>$ all the other numbers in (4.9) and this contradicts (4.8). Hence, $\phi_{81} = \phi_{71}$, which is then $>$ all the other numbers in (4.9), and in this case,

$$c_1 = a_3. \tag{4.11}$$

There remains now in (4.9) the following four ϕ_{jh}:

$$\phi_{61} = 1 - c_1 + c_3 > \phi_{62} = 1 - c_1,$$
$$\phi_{72} = a_1 - a_3 + c_3, \quad \phi_{82} = a_1,$$

of which $\phi_{61} > \phi_{62}$, and $\phi_{61} > \phi_{72}$ on account of (4.11) and (4.7). Therefore, by (4.8) and what we have proved above, the only possibility is that $\phi_{61} = \phi_{82}$, $\phi_{62} = \phi_{72}$, i.e.,

$$1 - c_1 + c_3 = a_1, \quad 1 - c_1 = a_1 - a_3 + c_3.$$

Combining these with (4.10) and (4.11), we get $a_1 = 3/4$, $a_3 = 1/2$, which contradict (4.7). Hence (4.6) cannot hold and Case 4 is impossible.

Lastly consider the remaining case.

Case 5. $\theta_6 = \theta_5 \neq$ any other θ_j.

In this case,

$$b_3 = a_1. \tag{4.12}$$

Moreover, by Lemma 4.2, we must have

$$u_5 + u_6 = ke((1 + a_1 - c_1)\gamma) - k(1 - w)e((a_1 + c_3)\gamma) - kwe((a_1 + c_3 - c_1)\gamma)$$
$$+ we((1 - c_1 + c_3)\gamma) - e((1 - c_1)\gamma) + (1 - w)e(c_3\gamma) \overset{\gamma}{\equiv} 0. \tag{4.13}$$

This identity implies, by (4.2) and Lemma 3.4, that first, the coefficients of γ of the different terms in (4.13) must be equal in pairs as follows:

$$1 + a_1 - c_1 = 1 - c_1 + c_3, \quad a_1 + c_3 = 1 - c_1, \quad a_1 + c_3 - c_1 = c_3; \tag{4.14}$$

and second, the sum of each corresponding pair of terms in (4.13) must be zero, i.e.,

$$k + w = 0, \quad k(1 - w) + 1 = 0, \quad -kw + 1 - w = 0. \tag{4.15}$$

It is easy to see that equations (4.14) are equivalent to

$$a_1 = c_1 = c_3 = \tfrac{1}{3}; \tag{4.16}$$

and equations (4.15), on account of (2.9) and (4.16), are equivalent to

$$w = -k = e(\pi/3). \tag{4.17}$$

Thus we have proved that, in this case, the parameters a_1, b_3, c_1, c_3 and w must have the values given in (4.12), (4.16) and (4.17); and we note by referring to (2.10) and (2.11) that these values are permissible values.

Return now to our main equation (3.1) ≡ (2.13). First, we see from (4.5) that the θ_j which can possibly be equal to $\theta_{12} = a_1 - b_1 - 1$ are

$$\theta_{13} = -b_1 + b_3 - 1, \quad \theta_{14} = a_1 - a_3 - 1,$$
$$\theta_{15} = -a_3 + b_3 - 1, \quad \theta_{17} = b_3 - 1.$$

But on account of (4.12) and (2.10), $\theta_{17} = a_1 - 1 \neq \theta_{12}$, and

$$\theta_{12} = a_1 - b_1 - 1 = \theta_{13}, \quad \theta_{14} = a_1 - a_3 - 1 = \theta_{15}. \tag{4.18}$$

Therefore, there are only the following two subcases to consider:

Case 5a. $\theta_{12} = \theta_{13} \neq$ any other θ_j.

Case 5b. $\theta_{12} = \theta_{13} = \theta_{14} = \theta_{15} \neq$ any other θ_j.

For Case 5a, Lemma 4.2 requires that

$$u_{12} + u_{13} = -k(1-w)e(a_1\gamma) + k(1-w)e((a_1 - c_1)\gamma)$$
$$+ (1-w) - (1-w)e(-c_1\gamma) \overset{\gamma}{\equiv} 0$$

which, on account of (4.16), reduces to

$$(1-w)\{-ke(\gamma/3) + (k+1) - e(-\gamma/3)\} \overset{\gamma}{\equiv} 0.$$

But this, by Lemma 3.2 and (4.17), cannot possibly hold. Therefore, Case 5a is impossible.

There remains then Case 5b, namely: $\theta_{12} = \theta_{13} = \theta_{14} = \theta_{15} \neq$ any other θ_j. In this case, we have from (4.18) that

$$b_1 = a_3; \tag{4.19}$$

and also (4.12), (4.16) and (4.17) hold. Moreover, by Lemma 4.2, we must have

$$\begin{aligned} u_{12} + u_{13} + u_{14} + u_{15} = & -k(1-w)e(a_1\gamma) + k(1-w)e((a_1 - c_1)\gamma) \\ & + (1-w) - (1-w)e(-c_1\gamma) \\ & - kle((a_1 - a_3 + c_3)\gamma) + kle((a_1 - a_3)\gamma) \\ & + le((c_3 - a_3)\gamma) - le(-a_3\gamma) \overset{\gamma}{\equiv} 0, \end{aligned}$$

which, on account of (4.16), reduces to

$$\begin{aligned} & -k(1-w)e(\gamma/3) + (k+1)(1-w)e(0\gamma) - (1-w)e(-\gamma/3) \\ & \quad - kle((\tfrac{2}{3} - a_3)\gamma) + (k+1)le((\tfrac{1}{3} - a_3)\gamma) - le(-a_3\gamma) \overset{\gamma}{\equiv} 0. \end{aligned} \tag{4.20}$$

Now, on account of (2.9) and (4.17), none of the six terms in (4.20) is zero. Therefore, by Lemma 3.4, the coefficients of γ in the first and fourth terms must be equal, i.e.,

$$a_3 = \tfrac{1}{3}; \tag{4.21}$$

and with this, (4.20) further reduces to

$$(1 + l - w)\{-ke(\gamma/3) + (k+1)e(0\gamma) - e(-\gamma/3)\} \overset{\gamma}{\equiv} 0.$$

But since $w = e(\pi/3)$ (cf. (4.17)) and $l = e(2a_3\pi) = e(2\pi/3)$ (cf. (2.9) and (4.21)), we have

$$1 + l - w = 1 + e(2\pi/3) - e(\pi/3) = 0.$$

Therefore, the identity above holds.

Now equations (4.12), (4.16), (4.17), (4.19) and (4.21) combined are equivalent to

$$a_1 = a_3 = b_1 = b_3 = c_1 = c_3 = 1/3, \quad w = e(\pi/3). \tag{4.22}$$

Since these values are permissible values of the parameters (cf. (2.10) and (2.11)), we can conclude from the discussions above that for equation (2.13) to become an identity in β, γ, it is necessary that (4.22) holds. And this means, in view of what was said at the end of §2, that the "only if" part of Theorem C is proved.

The "if" part of Theorem C is essentially Morley's theorem itself for which many proofs are known. However, we would get not only a new proof of Morley's theorem but also a check on our computation by verifying that equation (2.13) actually becomes an identity in β, γ when the parameters a_1, a_3, b_1, b_3, c_1, c_3 and w have the values given in (4.22). It turns out that we then have

$$\begin{aligned} &\theta_1 = \theta_2 = \theta_3 = \theta_4 = -\tfrac{1}{3},\\ &\theta_5 = \theta_6 = \tfrac{1}{3},\\ &\theta_7 = \theta_8 = \theta_9 = \theta_{10} = \theta_{11} = 0,\\ &\theta_{12} = \theta_{13} = \theta_{14} = \theta_{15} = -1,\\ &\theta_{16} = \theta_{17} = \theta_{18} = -\tfrac{2}{3}, \end{aligned} \tag{4.23}$$

and

$$w = e(\pi/3), \quad k = e(-2\pi/3), \quad l = e(2\pi/3);$$

and consequently, equation (2.13) may be rewritten as

$$\tilde{u}_1(\gamma)e(-\beta/3) + \tilde{u}_2(\gamma)e(\beta/3) + \tilde{u}_3(\gamma)e(0) + \tilde{u}_4(\gamma)e(-\beta) + \tilde{u}_5(\gamma)e(-2\beta/3) = 0. \tag{4.24}$$

Now, we can easily verify that the $\tilde{u}_\lambda(\gamma)$ in (4.24) are each a linear combination of $e(2\gamma/3)$, $e(\gamma)$, $e(\gamma/3)$, $e(0)$, $e(-\gamma/3)$ whose coefficients, on account of (4.23), are all equal to zero. Therefore, $\tilde{u}_\lambda(\gamma)$ are all identically zero in γ, and consequently, equation (2.13) $\equiv$ (4.24) is identically zero in β, γ. Theorem C is thus completely proved.

References

1. D. J. Kleven, Morley's theorem and a converse, this MONTHLY, 85 (1978) 100–105.
2. C. O. Oakley and J. C. Baker, The Morley trisector theorem, this MONTHLY, 85 (1978) 737–745.
3. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.
4. Y. C. Wong, Some properties of the triangle, this MONTHLY, 48 (1941) 530–535.
5. Y. C. Wong, Some extensions of the Douglas-Neumann theorem for concentric polygons, this MONTHLY, 75 (1968) 470–482.

ANSWERS TO "PHOTOS" ON PAGE 641

Above: Shizuo Kakutani; below: Claude Chevalley. Both pictures were taken in 1965.

THE FIXED POINT PROPERTY AND CARTESIAN PRODUCTS

ROBERT F. BROWN
Department of Mathematics, University of California, Los Angeles, CA 90024

1. INTRODUCTION

Kazimierz Kuratowski died in June, 1980, at the age of 84. Fifty years earlier he asked a question that puzzled topologists for a long time. This is the story of Kuratowski's question.

The Question and the Answer. A topological space has the fixed point property (fpp) if for any map (= continuous function) $f: X \to X$ there exists a fixed point, that is, a point $x \in X$ such that $f(x) = x$. Kuratowski's question was of the form: if spaces X and Y have the fpp, does their cartesian product $X \times Y$ have the fpp? (Recall that, as a set, $X \times Y$ consists of all ordered pairs (x, y) where $x \in X$ and $y \in Y$. The open sets in $X \times Y$ are the unions of cartesian products of open sets of X and of Y. In most of this paper, the topological setup is much simpler. Both X and Y will be subsets of euclidean spaces, so $X \times Y$ will just inherit its topology from a higher-dimensional euclidean space.)

At the risk of ruining all the suspense, I'll tell you right away that the answer to Kuratowski's question is "no," even if X and Y are required to be very well-behaved spaces.

Seeing the Answer. I will point out in the next section that we are discussing an old and rather basic problem, but one whose satisfactory resolution is quite recent. Although the complete solution involves some pretty sophisticated topology, you need remarkably little background information in order to understand the main points of the solution. The fact that the answer is negative helps explain why this should be so: the problem is solved by explicitly constructing spaces X and Y with the fpp such that $X \times Y$ lacks the fpp.

At times I will need to base my claims on advanced topics in topology. Even at these points it turns out to be easy to describe precisely what I'm using and to refer the curious reader to the relevant literature.

2. HISTORY

Kuratowski's Question. The published history of our problem began in 1930 when Kuratowski asked [**15**]: if X and Y are peano continua with the fpp, does $X \times Y$ have the fpp? (A *peano continuum* is a compact, connected, and locally connected metric space.)

I emphasize "published" history because I'm quite certain the problem was old and well known among topologists long before Kuratowski published it. Such a formal presentation usually indicates that a problem has been around long enough and has been discussed sufficiently so that the proposer is sure it is both difficult and of some significance to its subject area.

Furthermore, notice that Kuratowski put a hypothesis-peano-continuum on his spaces. This restriction suggests to me that enough was known about the question so that wildly pathological counterexamples had been discovered (though none seems to have been published at that time) or at least that the existence of such examples was suspected.

Motivation. Turning from the question of when the problem arose, let's ask a more important one: why would topologists be interested in it? Kuratowski didn't include any motivation for studying the problem, but I can suggest two reasons the problem came up.

Robert F. Brown received his Ph.D. from the University of Wisconsin where he was a student of Edward Fadell. He has since been at UCLA except for a visiting Professorship at the University of Arizona and two sabbatical years at the University of Warwick, England. His research area is algebraic topology, with a particular interest in fixed point theory.

The first reason is a concern with what might be called "the foundations of topology." The classical definition of topology is: the study of properties invariant under homeomorphisms. The fpp is such a property. What does it mean to "study" a property? The answer to that question would be very long, but certainly a part of the answer is: find out whether the property is preserved under the various basic constructions of topology—such as forming cartesian products. The most famous positive result, Tychonoff's Theorem, solves the cartesian product problem for the property of compactness. On the other hand, it has long been known that the cartesian product of normal spaces is not necessarily normal ([**19**] gives the standard example). Thus Kuratowski's question was, and is, interesting because its solution tells us something about the very nature of topology.

The second reason for being interested in the behavior of the fpp under cartesian products is tied up with the early history of topology. Let $\mathbf{R}^n$ denote euclidean n-dimensional space and let I^n be the standard n-cell, that is,

$$I^n = \{(x_1, x_2, \dots, x_n) \in \mathbf{R}^n \mid \; 0 \leqslant x_i \leqslant 1 \quad \text{for all} \quad i\}.$$

It is easy to show that $I = I^1$ has the fpp. The Brouwer Fixed Point Theorem states that I^n has the fpp for all n. Since I^n is the cartesian product of n copies of I, the Brouwer Theorem suggests that the fpp might behave well under the cartesian product construction. Moreover, although nowadays the Brouwer theorem is easy to prove using elementary algebraic topology, back in the 1920's the subject was much less well-developed and the existing proofs of the theorem were pretty difficult. But suppose the fpp *were* preserved under cartesian products. Then the Brouwer Fixed Point Theorem would be in immediate consequence: prove the easy $n = 1$ case and then apply induction.

A Pathological Example. It may be that Kuratowski's publication of the cartesian product problem for the fpp tended to focus topologists' attention on the problem. If it did, they don't seem to have had much success because there is no mention of the problem in the literature until many years later. A very special affirmative answer in 1956 by Eldon Dyer [**4**] (for "chainable continua") is probably best viewed as a generalization of the Brouwer Fixed Point Theorem.

The first significant contribution to the solution of Kuratowski's problem was the work of Edwin Connell. In 1959 he gave the first published example to demonstrate that without some restriction, such as Kuratowski's to peano continua, the fpp would not be preserved under cartesian product [**3**].

The example uses the subset X of $\mathbf{R}^2$ consisting of points $(x, \sin \frac{\pi}{1-x})$ for $0 \leqslant x < 1$ and the point $(1, 1)$, pictured in Fig. 1. It is easy to prove that X has the fpp. Connell then shows that $X \times X$ does not have the fpp. I won't repeat the argument, but I'll need to refer later to the method of proof. Connell proves that in any metric space with the fpp, "every locally finite chain of arcs is finite." (For the point I want to make, it doesn't matter what these words mean.) Then he constructs in $X \times X$ an infinite, locally finite chain of arcs. Thus, although all that is required to prove that a space lacks the fpp is to exhibit a single self-map of it without fixed points, it seems such a direct proof was not available for $X \times X$.

Kuratowski's problem was still unsolved because X is not a peano continuum—it is neither compact nor locally connected. But whether or not Kuratowski had been guided by a pathological example in stating the problem in 1930, such an example was now available.

Kuratowski's Question Answered. After 1959, other examples of spaces X and Y with the fpp such that $X \times Y$ lacks the fpp were published ([**13**] and [**14**]), but in neither case were both X and Y peano continua. Then, in 1967, Edward Fadell and his student William Lopez presented an example of a peano continuum (in fact a finite polyhedron) X with the fpp such that $X \times I$ doesn't have the fpp [**16**]. Thus Kuratowski's question was answered at last. I will describe such a polyhedron X in Section 4. It's not the original Fadell-Lopez example, but instead a somewhat simpler example suggested by Glen Bredon.

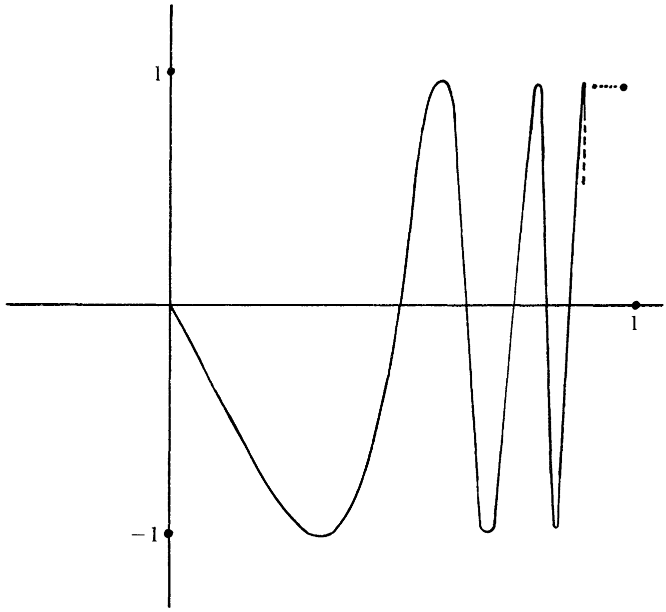


FIG. 1. Connell's example.

A Wedge is not Locally Euclidean. You have to know a little about these examples in order to understand what happened after 1967. The Fadell-Lopez type of example is a "wedge" of two sets. If X is a space (think of it as a subset of a euclidean space because that's what we'll be concerned with), A and B are closed subsets of X such that $X = A \cup B$ and $A \cap B = \{x_0\}$, a single point, then X is called the *wedge* of A and B. Write $X = A \vee B$. Notice that $X - \{x_0\}$ is disconnected.

The important property of a wedge for the purposes of fixed point theory is this easily proved fact: if A and B each have the fpp, so does $X = A \vee B$.

A space X is *locally n-euclidean* at a point x if there is a neighborhood U of x in X homeomorphic to $\mathbf{R}^n$. If a connected space X is locally n-euclidean at x, for $n \geqslant 2$, then $X - \{x\}$ is still connected. So if $X = A \vee B$ and $\{x_0\} = A \cap B$ then, since $X - \{x_0\}$ is disconnected, X certainly can't be locally n-euclidean at x_0, $n \geqslant 2$. Figure 2 shows just how very noneuclidean $X = A \vee B$ is at x_0, even when (as in this example), A and B are both locally euclidean at x_0.

The Manifold Problem. The Fadell-Lopez type of example is of the form $X = A \vee B$ where A and B are peano continua with the fpp, so X is a peano continuum and, as I pointed out above, it follows that X must also have the fpp. The observations that (1) his example depended on the wedge structure to prove X has the fpp and (2) a wedge is strikingly noneuclidean, at least at one point, led Fadell to wonder if the key to the example lay in the lack of locally euclidean structure. So, in 1970 [**5**], Fadell asked Kuratowski's question in this stronger form: if X and Y are compact manifolds with the fpp, does $X \times Y$ have the fpp? (An *n-manifold* is a metric space that is locally n-euclidean at every point.)

Almost immediately after Fadell raised the question, Bredon produced examples of nice spaces X and Y with the fpp such that $X \times Y$ does not have the fpp, and neither X nor Y is a wedge [**1**]. But Bredon's spaces are not manifolds according to the definition I just quoted. In 1977, Sufian Husseini [**11**] extended Bredon's ideas to construct examples that are manifolds, so the answer to Kuratowski's question is still "no" even if we require spaces as nice as these.

I will describe Husseini's manifolds in Section 5. One of the most agreeable features of the Bredon-Husseini work is the explicit construction of a map on the cartesian product that has no

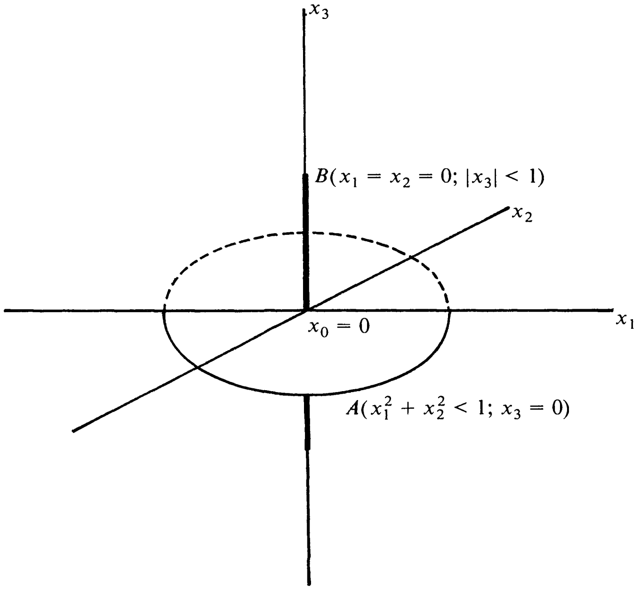


FIG. 2. The wedge.

fixed points. Remember that I said Connell's example $X \times X$ was shown by an indirect argument to lack the fpp. In the Fadell-Lopez type of example, the proof that $X \times I$ does not have the fpp comes from a general theory rather than anything intimately related to the construction of X. Thus, not only does Husseini's paper almost close the books on this part of fixed point theory, but it does it in a self-contained, constructive fashion, at least in comparison to the previous work.

You can skip the Fadell-Lopez-Bredon example by omitting Section 4 because Section 3 contains all the background you need for Section 5. On the other hand, if you just want to see an answer to Kuratowski's original question, the example in Section 4 is simpler than Husseini's material.

Unsolved Problems. I said that Husseini's paper almost completed the investigation of the behavior of the fpp when cartesian products are formed. In a way, that's not true. Any time the answer to a mathematical question is negative in the sense of "not always," people can ask under what additional hypotheses the answer becomes "yes"; though that's not necessarily an interesting question. When we ask—if X and Y are compact manifolds with the fpp, under what additional hypotheses is it true that the manifold $X \times Y$ has the fpp?—that *is* a worthwhile question. Relatively few manifolds with the fpp are known and without a wealth of examples it is difficult to study the property in this setting. Answers to the question should generate some of the examples we need. At present, very little is known about this problem; [**6**], [**12**] and [**25**] contain the only results published so far. I will discuss one of these positive theorems at the end of Section 4.

When we think only about the original, unadorned question: if X and Y have the fpp, does $X \times Y$ also, there really isn't much left to discuss. But Fadell has made one more observation. Recall that Connell's example was of a space X with the fpp for which $X \times X$ fails to have that property. The Fadell-Lopez type of example is of a peano continuum X with the fpp such that $X \times I$ lacks the fpp. It's not hard to show that the same X also produces an example where $X \times X$ fails to have the fpp. On the other hand, when in Section 5 I describe the construction of the map on the cartesian product of manifolds without fixed points, I will stress that it is crucial to the construction that the two manifolds are not homeomorphic. Thus Fadell has recently asked [**7**]: if X is a compact manifold with the fpp, does $X \times X$ have the fpp?

3. PRELIMINARIES

Quotient Spaces. Most of the constructions in this paper are of the following type. You have a topological space Y and an equivalence relation $\sim$ defined on Y. Write the equivalence class containing $y \in Y$ as $[y]$. Denote the set of equivalence classes by $Y/\sim$. Usually, instead of defining the equivalence relation $\sim$, it will be more convenient just to describe the nontrivial equivalence classes in $Y/\sim$, that is, the ones that consist of more than one point of Y. Define the function $q: Y \to Y/\sim$ by $q(y) = [y]$. The *quotient topology* on $Y/\sim$ is defined by: $U \subseteq Y/\sim$ is an open set if and only if $q^{-1}(U)$ is an open subset of Y. The set $Y/\sim$ will always be equipped with this topology and so it is called a *quotient space*.

The Suspension. For our first construction, suppose A is a space and let $[-1, 1] = \{x \in \mathbf{R}^1 \mid -1 \leqslant x \leqslant 1\}$. Choose the nontrivial equivalence classes of $(A \times [-1, 1])/\sim$ to be $A \times \{-1\}$, which I'll abbreviate as $\mathbf{-1}$, and $\mathbf{1} = A \times \{1\}$. The quotient space $(A \times [-1, 1])/\sim$ is called the *suspension* of A and written ΣA (see Fig. 3).

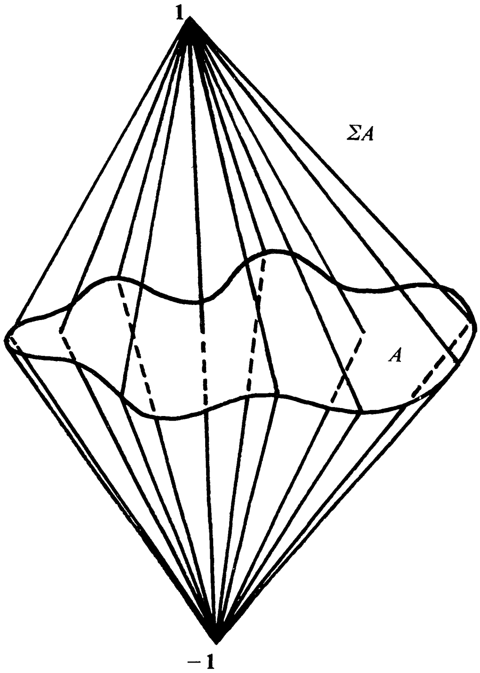


FIG. 3. The suspension.

It will at times be convenient to consider the sphere S^{n-1} as the boundary of the cell I^n that I defined before. That is, let

$$S^{n-1} = \{x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \mid 0 \leqslant x_i \leqslant 1 \text{ for all } i \text{ and } x_i = 0 \text{ or } x_i = 1 \text{ for at least one } i\}.$$

It is not hard to see that the suspension of a sphere is homeomorphic to a sphere, in symbols $\Sigma S^{n-1} = S^n$.

Spherical Complexes. All the examples make use of a basic geometric construction called "attaching a cell." Start with a space A and a map $g: S^{n-1} \to A$. I will attach I^n to A *via* the map g to form a new space denoted by $A \cup_g I^n$. The set $A \cup I^n$ is topologized by declaring a subset open if its intersections with A and I^n are both open. This space is called the "disjoint union" of A

and I^n and written $A \amalg I^n$. The space $A \cup_g I^n$ is $(A \amalg I^n)/\sim$ where the nontrivial equivalence classes are of the form $\{a\} \cup \{g^{-1}(a)\}$ for all a in $g(S^{n-1})$.

If A is a subset of a euclidean space $\mathbf{R}^m$, it is not difficult to construct a subset of $\mathbf{R}^{m+n+1}$ homeomorphic to $A \cup_g I^n$. This construction is easier to visualize than the quotient space definition of $A \cup_g I^n$, and I will use the construction in Section 5, so let me briefly indicate how it goes. Think of $\mathbf{R}^{m+n+1}$ as $\mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R}^1$ and denote the origin in a euclidean space of dimension greater than one by $\mathbf{0}$. Put A in $\mathbf{R}^m \times \{\mathbf{0}\} \times \{1\}$ and I^n in $\{\mathbf{0}\} \times \mathbf{R}^n \times \{0\}$. Instead of forming equivalence classes, for each x in s^{n-1} join x to $g(x)$ by a line segment in $\mathbf{R}^{m+n+1}$. The union of A, I^n, and all the line segments, is homeomorphic to $A \cup_g I^n$ (see Fig. 4).

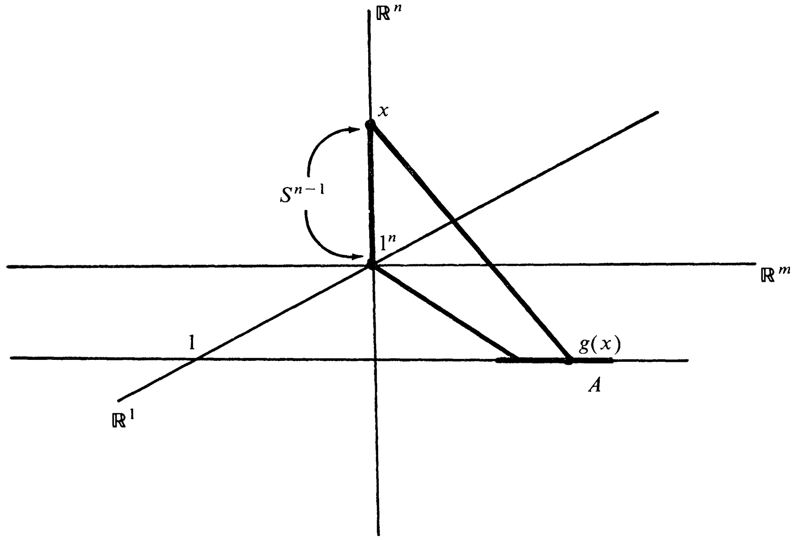


FIG. 4. Attaching a cell.

Notice that if A is a point, so $g: S^{n-1} \to A$ is the constant map, then $A \cup_g I^n$ is homeomorphic to S^n. (Imagine shrinking A to a point in Fig. 4, so $A \cup_g I^n$ is a triangle, or look ahead to Fig. 7.)

Now I'll use the attaching-a-cell construction to build a class of spaces. Begin with a point and attach a cell. Assuming we have a map from a sphere (of any dimension) to the set we just constructed (a sphere at this stage), attach a cell. Continuing in this way—attaching cells one at a time—a finite number of times, we end up with a compact set. A space constructed in this way is called a *spherical complex*. (Section 19 of [**8**] is a good place to start on a more systematic study of spherical complexes.)

Complex Projective Spaces. Identify the complex numbers $\mathbf{C}$ with the euclidean plane $\mathbf{R}^2$ ($z = x + iy$ becomes (x, y)) and, more generally, complex n-space $\mathbf{C}^n = \mathbf{C} \times \mathbf{C} \times \cdots \times \mathbf{C}$ with $\mathbf{R}^{2n}$. I will also think of $\mathbf{C}^n$ as an n-dimensional vector space with vectors $z = (z_1, z_2, \ldots, z_n)$. The norm of z,

$$\|z\| = (z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots + z_n\bar{z}_n)^{1/2}$$

($\bar{z}_i$ is the complex conjugate of z_i) measures distance to the origin in $\mathbf{R}^{2n}$, so we can define sets homeomorphic to odd-dimensional spheres by $S^{2n-1} = \{z \in \mathbf{C}^n \mid \|z\| = 1\}$.

The complex projective spaces are defined by an equivalence relation on odd-dimensional spheres. Call $z, z' \in S^{2n+1}$ (in $\mathbf{C}^{n+1}$) equivalent, and write $z \sim z'$, if there exists $\lambda \in \mathbf{C}$ such that $z' = \lambda z$ (scalar multiplication in $\mathbf{C}^{n+1}$). The space $S^{2n+1}/\sim$ is *complex projective n-space*, denoted by $\mathbf{C}P^n$.

The complex projective spaces are spherical complexes. Starting with the number $1 \in \mathbf{R}^1$, there are maps $g_1, g_2, \ldots, g_n$ such that $1 \cup_{g_1} I^2 \cup_{g_2} I^4 \cup \cdots \cup_{g_n} I^{2n}$ is homeomorphic to $\mathbf{C}P^n$. The details of the construction can be found in [**10**, pp. 138–142]. You don't have to know how the maps g_k are defined to understand the rest of this paper, so I won't go into it here.

I'll use the existence of the spherical complex structure to establish a property of $\mathbf{C}P^n$ that will be very important later: $\mathbf{C}P^n$ is a $2n$-manifold. The last step in the spherical complex construction of $\mathbf{C}P^n$ is attaching the cell I^{2n}. Let $[z^*] \in \mathbf{C}P^n$ be a point in the interior of I^{2n}, then $\mathbf{C}P^n$ is locally $2n$-euclidean at $[z^*]$ because the interior of I^{2n} is homeomorphic to $\mathbf{R}^{2n}$. Now take any point $[z'] \in \mathbf{C}P^n$. Think of $\mathbf{C}P^n$ as $S^{2n+1}/\sim$ and let $z^*, z' \in S^{2n+1} \subset \mathbf{C}^{n+1}$ represent the corresponding equivalence classes. There is a nonsingular linear transformation $T: \mathbf{C}^{n+1} \to \mathbf{C}^{n+1}$ such that $T(z^*) = z'$. Defining $[T]: \mathbf{C}P^n \to \mathbf{C}P^n$ by $[T][z] = [T(z)]$ produces a homeomorphism taking $[z^*]$ to $[z']$. The neighborhood of $[z^*]$ homeomorphic to $\mathbf{R}^{2n}$ is carried onto a neighborhood of $[z']$ and I've shown that $\mathbf{C}P^n$ is locally $2n$-euclidean at $[z']$.

Lefschetz Theory. All the subsets of euclidean spaces that I will describe later are very well-behaved spaces from the point of view of algebraic topology. This subject supplies the tools for proving that spaces have the fpp. The very brief excursion into algebraic topology that follows is intended to give you a general idea of the approach. If you want to know more, you might start with [**2**].

Let X be one of these well-behaved spaces and let $\mathbf{K}$ be a field. The fields we will use are $\mathbf{Q}$, the rationals, and finite fields $\mathbf{Z}/p$, p a prime. There is a set of finite dimensional $\mathbf{K}$-vector spaces $H_0(X; \mathbf{K}), H_1(X; \mathbf{K}), \ldots, H_m(X; \mathbf{K})$, the homology of X with $\mathbf{K}$ coefficients. A map $f: X \to X$ gives rise to a set of linear transformations $H_i(f; \mathbf{K}): H_i(X; \mathbf{K}) \to H_i(X; \mathbf{K})$ for $i = 0, 1, \ldots, m$. Each linear transformation has a trace, denoted by $\operatorname{Tr}[H_i(f; \mathbf{K})]$. The $\mathbf{K}$-*Lefschetz number* $L(f; \mathbf{K})$ of f is

$$L(f; \mathbf{K}) = \sum_{i=0}^{m} (-1)^i \operatorname{Tr}\left[H_i(f; \mathbf{K})\right].$$

The Lefschetz Fixed Point Theorem, which holds for nice spaces like the ones we are interested in, states that if $L(f; \mathbf{K}) \neq 0$ for any field $\mathbf{K}$, then f has a fixed point.

A space X has *property* $F(\mathbf{K})$ (see [**6**]) if $L(f; \mathbf{K}) \neq 0$ for all maps $f: X \to X$. If the Lefschetz Theorem applies to the space X, then property $F(\mathbf{K})$, for any field $\mathbf{K}$, implies the fpp.

To apply the Lefschetz theory to prove that a space X has the fpp, you first have to know a lot about the homology of X. Often this knowledge comes from information about how the space is constructed, for instance as a spherical complex. Now a vector space certainly has a lot of linear transformations defined on it. But the linear transformations $H_i(f; \mathbf{K})$ are very special because they are obtained from maps. Linear transformations of homology that come from maps have many algebraic properties not shared by the general run of linear transformations. So it's not surprising that the possible values of $L(f; \mathbf{K})$, which is defined algebraically from the $H_i(f; \mathbf{K})$, may be very much restricted because of the specialness of the $H_i(f; \mathbf{K})$. A careful analysis of what is possible for the particular space X may lead to the conclusion that for any map $f: X \to X$ the $H_i(f; \mathbf{K})$ are such that $L(f; \mathbf{K}) \neq 0$. So X has property $F(\mathbf{K})$ and therefore the fpp.

4. SOLUTION TO KURATOWSKI'S PROBLEM

Quaternionic Projective Spaces. Let's think of the complex numbers $\mathbf{C}$ as a 2-dimensional real vector space with a multiplication

$$(x, y)(x', y') = (xx' - yy', xy' + x'y).$$

Then we're prepared to view the set of quaternions, denoted by $\mathbf{H}$, as a 4-dimensional real vector space with a similar but more complicated multiplication:

$$(w, x, y, z)(w', x', y', z') = (ww' - xx' - yy' - zz',$$
$$wx' + xw' + yz' - zy', wy' + yw' - xz' + zx', wz' + zw' + xy' - yx').$$

Naturally, we can identify $\mathbf{H}$ with $\mathbf{R}^4$ and quaternionic n-space $\mathbf{H}^n = \mathbf{H} \times \mathbf{H} \times \cdots \times \mathbf{H}$ with $\mathbf{R}^{4n}$. The conjugate of a quaternion $q = (w, x, y, z)$ is $\bar{q} = (w, -x, -y, -z)$ so the norm of $q = (q_1, q_2, \ldots, q_n) \in \mathbf{H}^n$, that is

$$\|q\| = (q_1\bar{q}_1 + q_2\bar{q}_2 + \cdots + q_n\bar{q}_n)^{1/2}$$

is the distance to the origin in $\mathbf{R}^{4n}$ and we can define spheres

$$S^{4n-1} = \{q \in \mathbf{H}^n \mid \|q\| = 1\}.$$

Quaternionic projective spaces are the analogue of complex projective spaces. Define an equivalence relation on $S^{4n+3} \subset \mathbf{H}^{n+1}$ by setting $q \sim q'$ if there exists $\lambda \in \mathbf{H}$ such that $q' = \lambda q$. Then $S^{4n+3}/\sim$ is *quaternionic projective n-space*, denoted by $\mathbf{H}P^n$. The space $\mathbf{H}P^n$ has the structure of a spherical complex of the form

$$1 \cup_{k_1} I^4 \cup_{k_2} I^8 \cup \cdots \cup_{k_n} I^{4n},$$

and it is a $4n$-manifold. Lefschetz theory can be used to prove that $\mathbf{H}P^n$ has property $F(\mathbf{Z}/3)$ for all $n \geqslant 2$. A sketch of the proof can be found in pages 21–22 of [**5**].

The Example. Complex and quaternionic projective spaces are the building blocks of the peano continuum X that answers Kuratowski's question because X has the fpp but $X \times I$ does not.

I said in Section 2 that X is a wedge of spaces. Let's see how to construct a wedge. Choose points $x_0 \in A$ and $y_0 \in B$. Imbed A and B in $A \times B$ by sending $x \in A$ to (x, y_0) and $y \in B$ to (x_0, y). (Compare Figure 2 where $A \subset \mathbf{R}^2$, $B \subset \mathbf{R}^1$, and x_0, y_0 are the origins of the respective euclidean spaces.) The union of the imbedded sets I will call $A \vee B$ because it really is the wedge as defined in Section 2.

A space that answers Kuratowski's question is

$$X = \mathbf{H}P^3 \vee \Sigma\mathbf{C}P^4.$$

Certainly X is a very nice peano continuum. I have already remarked that $\mathbf{H}P^3$ has the fpp, because it has property $F(\mathbf{Z}/3)$. The same theory implies that $\Sigma\mathbf{C}P^n$ has property $F(\mathbf{Z}/2)$ provided that n is even, so $\Sigma\mathbf{C}P^4$ has the fpp also. (There is an indication of the proof in [**5**], on page 18.) As I pointed out in Section 2, it follows that the wedge X has the fpp.

Applying Wecken's Theorem. The space $X \times I$ doesn't have the fpp because it satisfies the hypotheses of a general theorem on the existence of fixed point free maps (= maps without fixed points). The theorem has two hypotheses, one algebraic and one geometric. I'll first explain the hypotheses and state the theorem, and then I'll indicate how to show that the theorem does apply to $X \times I$.

The algebraic hypothesis uses the field $\mathbf{Q}$ of rational numbers. Continue to assume the spaces are nice enough so that Lefschetz theory can be used. Let Y be such a space and denote by 1_Y the identity map on Y: $1_Y(y) = y$ for all $y \in Y$. The *Euler characteristic* of Y, written $\chi(Y)$, is defined to be the $\mathbf{Q}$-Lefschetz number of 1_Y; in symbols $\chi(Y) = L(1; \mathbf{Q})$. To compute $\chi(Y)$ you just have to know the homology $H_i(Y; \mathbf{Q})$, $i = 0, 1, \ldots, m$. The reason is that for any field $\mathbf{K}$, $H_i(1_Y; \mathbf{K})$ is the identity transformation of $H_i(Y; \mathbf{K})$ so the trace $\mathrm{Tr}[H_i(1_Y; \mathbf{K})]$ is just the dimension of the vector space $H_i(Y; \mathbf{K})$. Therefore, to compute the Euler characteristic, use

$$\chi(Y) = \sum_{i=0}^{m} (-1)^i \mathrm{Dim}[H_i(Y; \mathbf{Q})].$$

If $\chi(Y) \neq 0$, then by the Lefschetz Fixed Point Theorem any map $f: Y \to Y$ that induces the identity transformation on all $H_i(Y; \mathbf{Q})$ has a fixed point because then $L(f; \mathbf{Q}) = \chi(Y)$. In particular, any map homotopic to 1_Y has that property and consequently it has a fixed point. The converse of that last statement says that if $\chi(Y) = 0$, then some map homotopic to 1_Y is fixed point free. The converse is *not* true in general, but it is if strong enough topological conditions are imposed on Y. For instance, Hopf proved in 1927 [**9**] that if Y is a compact differentiable manifold and $\chi(Y) = 0$, then Y admits a nonsingular vector field, and it is an easy consequence that such a manifold Y admits a fixed point free map $f: Y \to Y$ homotopic to 1_Y.

At the end of this section, I will outline the proof that, for the example $X = \mathbf{H}P^3 \vee \Sigma\mathbf{C}P^4$, we have $\chi(X \times I) = 0$. Hopf's theorem will not produce a fixed point free map of $X \times I$, however, because $X \times I$ is not a manifold, but a theorem of Franz Wecken will. To state Wecken's theorem, I need the following definition. A point y in a space Y is called a *local cut point* if there exists a connected neighborhood U of y such that $U - \{y\}$ is disconnected. The left-hand portion of Fig. 5 shows a space with a local cut point at x^*. Back in Section 2, I pointed out that manifolds don't have local cut points but that wedges definitely do have them.

Now I can state the theorem, published by Wecken in 1942 [**24**] (also, see [**2**]): If Y is a finite polyhedron with no local cut point and $\chi(Y) = 0$, then there is a fixed point free map $f: Y \to Y$ homotopic to 1_Y. The space $X \times I$ is such a polyhedron, it has no local cut points, and $\chi(X \times I) = 0$, so $X \times I$ does not have the fpp. In the rest of this section, I'll sketch the arguments that allow Wecken's theorem to be applied to $X \times I$.

The complex and quaternionic projective spaces are finite polyhedra. One way to prove this is to show that each time a cell is attached the new space can be triangulated because the map from the sphere that does the attaching is so well-behaved. Alternatively, the projective spaces are actually differentiable manifolds, and these are known to be polyhedra, see [**17**]. The suspension of a finite polyhedron is a finite polyhedron. If you form the wedge of two finite polyhedra so that their intersection is a vertex of each, it's evident that you still have a finite polyhedron. So X is a finite polyhedron and since cartesian products of finite polyhedra also have that structure, $X \times I$ is a finite polyhedron, as I claimed.

Since $X = \mathbf{H}P^3 \vee \Sigma\mathbf{C}P^4$ does have a local cut point, I surely need to explain why $X \times I$ doesn't have one. The proof depends on the fact that X is a finite polyhedron and therefore locally pathwise-connected. (The space X is *locally pathwise-connected* if for any $x \in X$ and neighborhood U of x there is a pathwise-connected neighborhood V of x in U. That is, for $x' \in V$ there is a path in V from x' to x.) Fig. 5 and 6 indicate why taking the cartesian product with I eliminates local cut points.

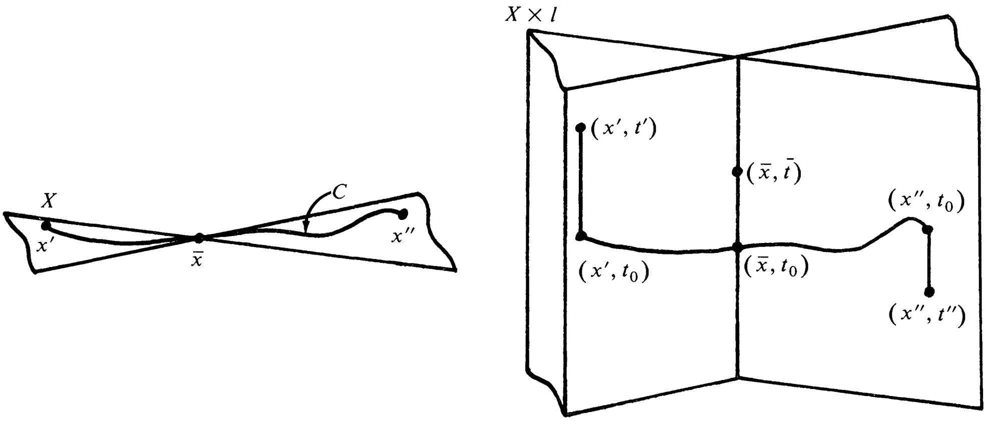


FIG. 5. Local cut points — case 1.

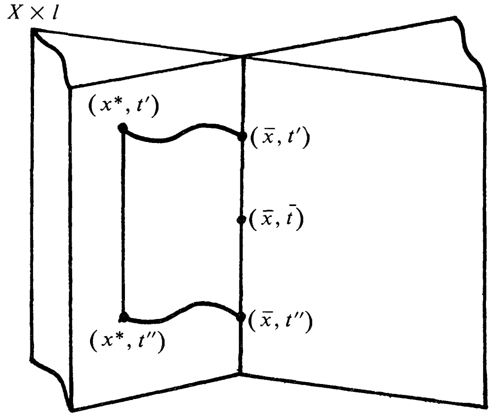


FIG. 6. Local cut points — case 2.

Here is how it works. Suppose we are given $(\bar{x}, \bar{t}) \in X \times I$ and a neighborhood W of it. Figs. 5 and 6 show how to find a neighborhood U of $(\bar{x}, \bar{t})$ in W such that $U - \{(\bar{x}, \bar{t})\}$ is pathwise-connected and therefore connected. To do it, we use the local pathwise-connectedness of X to find a pathwise-connected neighborhood V of $\bar{x}$ in X such that $U = V \times I_0$ is contained in W, where I_0 is an open interval in I containing $\bar{t}$. Given (x', t') and (x'', t'') in $U - \{(\bar{x}, \bar{t})\}$ with x' and x'' not both $\bar{x}$, choose $t_0 \neq \bar{t}$ in I_0 and define a path between them as in Fig. 5. The path C in V from x' to x'' exists because V is pathwise-connected. If $x' = x'' = \bar{x}$, pick some $x^* \in V$, let C now be a path in V from $\bar{x}$ to x^* and proceed as in Fig. 6.

Figs. 5 and 6 illustrate the case where $\bar{x}$ is a local cut point, but the argument applies as well to any point of X.

The last thing I have to do is calculate the Euler characteristic of $X \times I$. From the homology I know that $\chi(\mathbf{H}P^n) = n + 1$ and $\chi(\Sigma\mathbf{C}P^n) = 1 - n$, for all n. The Euler characteristics of wedges and cartesian products behave according to the equations

$$\chi(A \vee B) = \chi(A) + \chi(B) - 1$$
$$\chi(X \times Y) = \chi(X)\chi(Y).$$

The calculation then goes:

$$\begin{aligned}\chi(X \times I) &= \chi(X)\chi(I) = \chi(\mathbf{H}P^3 \vee \Sigma\mathbf{C}P^4)\chi(I)\\ &= \left[\chi(\mathbf{H}P^3) + \chi(\Sigma\mathbf{C}P^4) - 1\right]\chi(I)\\ &= [(4) + (-3) - 1]\chi(I) = 0.\end{aligned}$$

This calculation depends only on relatively elementary topics from homology theory: homology of projective spaces, the Mayer-Vietoris sequence, and the Künneth Theorem, so a book like [**23**] is a good reference if you want to fill in details.

Now we know the answer to Kuratowski's question is "no," but it seems a pity to end the section in such a discouraging way. So I'll describe a positive result concerning the fixed point property and cartesian products.

First we need another version of Wecken's Theorem [**24**], for simply-connected polyhedra. (Remember that a space X is *simply-connected* if any map $g: S^1 \to X$ can be extended to a map $G: I^2 \to X$.) Wecken proved that if X is a simply-connected finite polyhedron with no local cut points and $f: X \to X$ is a map such that $L(f; \mathbf{Q}) = 0$, then f is homotopic to a fixed point free map. Therefore, such a polyhedron X can have the fpp only if X has property $F(\mathbf{Q})$.

The other ingredient we require is a theorem of Fadell. Suppose X and Y are finite polyhedra

with *disjoint rational homology*, that is, $H_i(X; \mathbf{Q}) \neq 0$ for some $i \geqslant 1$ implies $H_i(Y; \mathbf{Q}) = 0$. The very helpful referee informs me that a small modification of the proof of Theorem 4.2 of [**6**] establishes the fact that if such X and Y both have property $F(\mathbf{Q})$, then $X \times Y$ also has property $F(\mathbf{Q})$.

Combining Wecken's and Fadell's results, we have:

THEOREM. *Let X and Y be simply-connected finite polyhedra with no local cut points and disjoint rational homology. If X and Y have the fpp, then $X \times Y$ also has the fpp.*

Examples of polyhedra satisfying the hypotheses of the Theorem are $X = \mathbf{C}P^{2n}$ and $Y = \Sigma \mathbf{C}P^{2m}$, for $m, n \geqslant 1$, so it implies that $\mathbf{C}P^{2n} \times \Sigma \mathbf{C}P^{2m}$ has the fpp.

5. THE MANIFOLD EXAMPLE

Maps on spheres. Recall from Section 3 that we can think of the sphere S^n as the spherical complex $1 \cup_c I^n$. Use that structure to define a map $k : I^n \to S^n$ in the following way: collapse the boundary of I^n to 1 and wrap the interior of I^n homeomorphically around $S^n - \{1\}$ (see Fig. 7, where $[q, r]$ denotes the line segment between q and r). For any space X, there is a natural

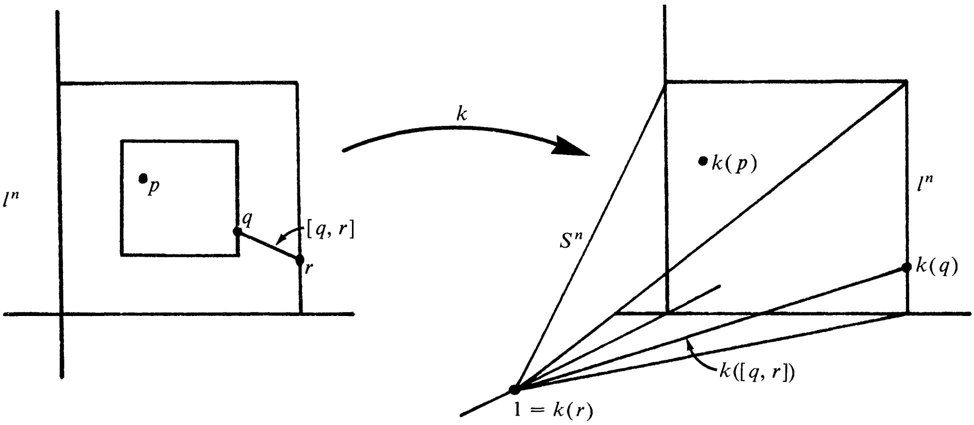


FIG. 7. Mapping cell to sphere.

one-to-one correspondence between all maps from S^n to X and those maps from I^n to X which are constant on the boundary of I^n. To describe the correspondence, let $a : S^n \to X$ be any map, then $a' = ak : I^n \to X$ has the required property. The inverse correspondence works like this. Let $a' : I^n \to X$ be a map taking the boundary of I^n to $x_0 \in X$. Define $a : S^n \to X$ by setting $a(p) = a'k^{-1}(p)$ if $p \neq 1$ and $a(1) = x_0$.

Given maps $a', b' : I^n \to X$ that are constant on the boundary of I^n, we can "add" them to form a new map $a' + b' : I^n \to X$ by splitting I^n in half and letting $a' + b'$ equal a' on the bottom and b' on the top. Formally, for $p = (p_1, p_2, \ldots, p_n)$ in I^n, let

$$(a' + b')(p) = \begin{cases} a'(p_1, p_2, \ldots, p_{n-1}, 2p_n) & 0 \leqslant p_n \leqslant 1/2 \\ b'(p_1, p_2, \ldots, p_{n-1}, 2p_n - 1) & 1/2 \leqslant p_n \leqslant 1. \end{cases}$$

The fact that a' and b' are constant on the boundary of I^n makes $a' + b'$ a continuous function with the same property.

For $a' : I^n \to X$ as before and some positive integer N, we can use successive additions to produce a map $Na' : I^n \to X$. The definition carries over to maps on S^n: for $a : S^n \to X$, corresponding to $a' : I^n \to X$, let $Na : S^n \to X$ be the map corresponding to Na'.

Although a full understanding of the homotopy groups of the spheres remains beyond the

reach of present-day algebraic topology, much has been discovered about them. The following fact comes from that body of knowledge [22]. There is a map $\phi: S^{23} \to S^{13}$ such that 6ϕ is homotopic to the constant map, but for $N = 1, 2, \ldots, 5$ the map $N\phi$ is not homotopic to the constant map. The map ϕ is the key to Husseini's manifolds with the fpp whose cartesian product fails to have the fpp.

Building the Manifolds. I will use the maps $2\phi, 3\phi: S^{23} \to S^{13}$ to construct two manifolds. Since the steps are identical, except for the choice of map to start with, I'll just describe one of the manifolds.

We will form the spherical complex $C(2\phi) = 1 \cup_c I^{13} \cup_{2\phi} I^{24}$. The map $c: S^{12} \to \{1\}$ is of course constant. If we think of the point $\{1\}$ as $\mathbf{R}^0$, then the construction I described in Section 3 places $1 \cup_c I^{13}$ in $\mathbf{R}^{0+13+1} = \mathbf{R}^{14}$ (compare the right hand side of Fig. 7). Then attaching the cell I^{24} by means of that same construction, we have built $C(2\phi) = (1 \cup_c I^{13}) \cup_{2\phi} I^{24}$ in $\mathbf{R}^{14+24+1} = \mathbf{R}^{39}$. Writing $\mathbf{R}^{56} = \mathbf{R}^{39} \times \mathbf{R}^{17}$, we can view $C(2\phi)$ as a subset of $\mathbf{R}^{56}$ by letting $x \in C(2\phi)$ correspond to $(x, \mathbf{0})$. It's all right to assume that the map ϕ is very well behaved. If it isn't, I'll homotope it to a better form because changing a map by a homotopy doesn't affect whether multiples of it are homotopic to the constant map. Consequently, I can see to it that $C(2\phi)$ is nice enough so that it has a *regular neighborhood* $R(2\phi)$ in $\mathbf{R}^{56}$. That is to say, there is a number $\varepsilon > 0$ with the following property: For a point $p \in \mathbf{R}^{56}$, let $d(p, C(2\phi))$ denote the distance from p to $C(2\phi)$ and let

$$R(2\phi) = \{p \in \mathbf{R}^{56} \mid d(p, C(2\phi)) \leqslant \varepsilon\}$$

Then there is a retraction $\rho: R(2\phi) \to C(2\phi)$. (Recall that if $A \subseteq X$, then a *retraction* of X onto A is a map $g: X \to A$ such that $g(a) = a$ for all $a \in A$.)

If $d(p, C(2\phi)) < \varepsilon$, then p has a neighborhood in $\mathbf{R}^{56}$ with the same property, so $R(2\phi)$ is locally 56-euclidean at such a point. But $R(2\phi)$ is not locally euclidean at p if $d(p, C(2\phi)) = \varepsilon$. However, there's still a good structure there. Let

$$\mathbf{R}^n_+ = \{(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n \mid x_n \geqslant 0\}$$

and call a space S an *n-manifold with boundary* if each $s \in S$ has a neighborhood homeomorphic either to $\mathbf{R}^n$ or to $\mathbf{R}^n_+$. The set of points of S with neighborhoods homeomorphic to $\mathbf{R}^n_+$ is called the *boundary* of S and denoted by ∂S. The phrase "with boundary" will imply $\partial S \neq \varnothing$ and I'll reserve the term "n-manifold" for locally euclidean spaces as before. Now $R(2\phi)$ is a 56-manifold with boundary, where $\partial R(2\phi) = \{p \in \mathbf{R}^{56} \mid d(p, C(2\phi)) = \varepsilon\}$. To learn more about regular neighborhoods, see [**18**].

Let R be an n-manifold with boundary. It doesn't have to be $R(2\phi)$, though that's what I have in mind. I want to imbed R in an n-manifold, and have a retraction of the manifold back to R. To do it, take two identical copies of R. We can eliminate some potential confusion by calling them R and R'. Then, for each x in R, the same point in R' will be called x'. Define an equivalence relation on the disjoint union $R \amalg R'$ so that the nontrivial equivalence classes are all the pairs $\{x, x'\}$ for x in ∂R. The quotient space $(R \amalg R')/\sim$ is called the *double* of R and denoted by $2R$. Away from the nontrivial equivalence classes, $2R$ is homeomorphic to R and therefore locally n-euclidean. A neighborhood of $[x, x']$ in $2R$ is homeomorphic to two copies of $\mathbf{R}^n_+$, which looks like half of $\mathbf{R}^n$, joined together to form a complete copy of $\mathbf{R}^n$ itself. So $2R$ is an n-manifold (see Fig. 8 where R is homeomorphic to I^2 and, joining two copies along the boundary S^1, we get a space $2R$ homeomorphic to S^2). The retraction $\pi: 2R \to R$ is defined by

$$\pi[x] = \pi[x, x'] = \pi[x'] = x$$

which just amounts to pushing the R' portion of $2R$ down onto the R portion.

Forming $2R(2\phi)$ produces a 56-manifold, but it takes one more step to get a manifold that can be shown by Lefschetz theory to have the fpp. The final step uses the connected sum construction

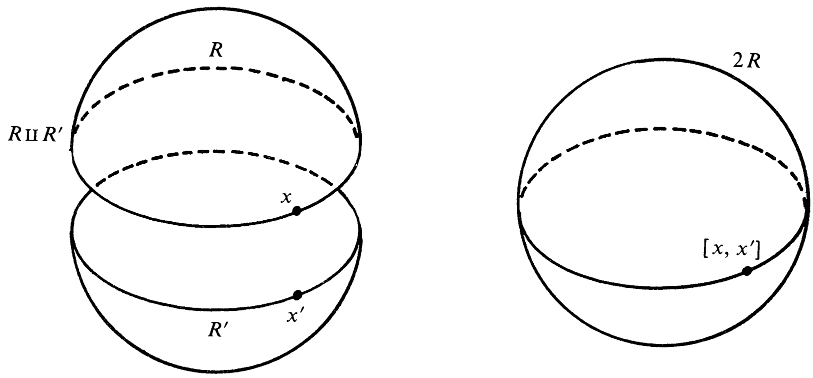


FIG. 8. The double.

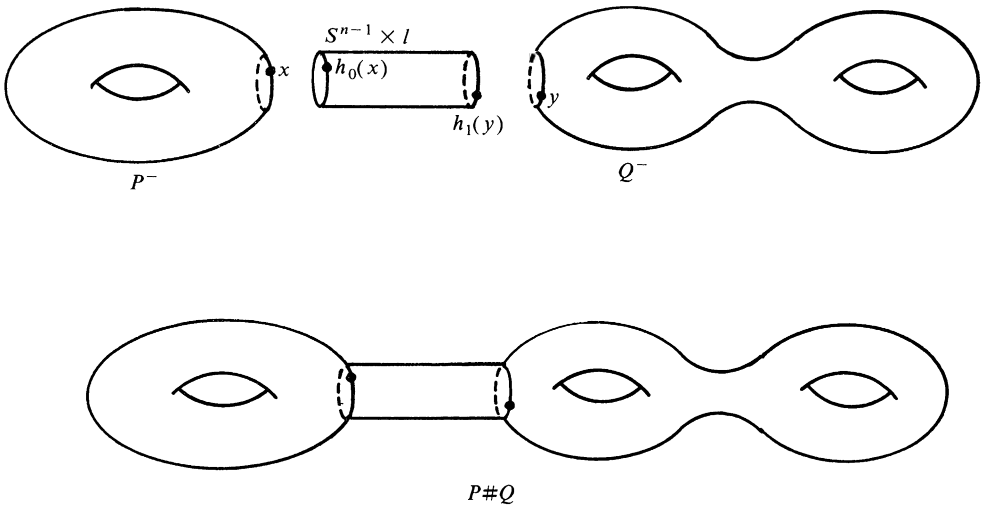


FIG. 9. The connected sum.

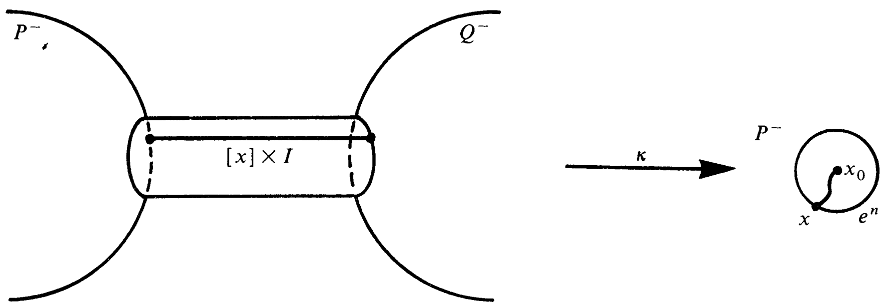


FIG. 10. The map κ.

to join two disjoint n-manifolds into a single n-manifold. Suppose P and Q are n-manifolds. In each of P and Q find a subset homeomorphic to I^n, remove the interiors, and call the resulting n-manifolds with boundary P^- and Q^-. The boundaries of P^- and Q^- are homeomorphic to S^{n-1}. I will join P^- and Q^- by a "tube" $S^{n-1} \times I$. Let

$$h_0 : \partial P^- \to S^{n-1} \times \{0\} \text{ and } h_1 : \partial Q^- \to S^{n-1} \times \{1\}$$

be homeomorphisms. Form the disjoint union $P^- \amalg S^{n-1} \times I \amalg Q^-$ and take as nontrivial equivalence classes $\{x, h_0(x)\}$ for all $x \in \partial P^-$ and $\{y, h_1(y)\}$ for all $y \in \partial Q^-$. The quotient space is called the *connected sum* of the manifolds P and Q; it is denoted by $P\#Q$.

I outlined an argument to show that joining the n-manifolds with boundary R and R' to form $2R$ produces an n-manifold. The same argument proves that when the three n-manifolds with boundary P^-, $S^{n-1} \times I$, and Q^- are joined to form $P\#Q$, an n-manifold is again the result. Fig. 9 illustrates the connected sum construction for the 2-manifolds P, the torus, and Q, the two-holed torus.

Next I'll define a map from $P\#Q$ into P. It's not quite a retraction; after all, P isn't even a subset of $P\#Q$. But the map is the identity on P, and that will turn out to be the property I need. Write $P = P^- \cup e^n$ where e^n is homeomorphic to I^n and notice that $\partial P^- = \partial e^n$. Take a point x_0 in the interior of e^n, and let's think of e^n as a union of arcs (homeomorphic images of I) from ∂e^n to x_0. Use the symbol $\langle x, x_0 \rangle$ for such an arc, $x \in \partial e^n$. Then look at Fig. 10 to picture the following description of $\kappa : P\#Q \to P$. Let $\kappa(x) = x$ for all $x \in P^-$, let $\kappa(y) = x_0$ for all $y \in Q^-$, and then require that, for $x \in \partial P^-$, the map κ take $[x] \times I$ homeomorphically onto $\langle x, x_0 \rangle$ in a way that varies continuously with x. (It's not hard to do; I'll skip the formal definition.)

In Section 3, we saw that the complex projective space $\mathbf{C}P^n$ is a $2n$-manifold. Now $2R(2\phi)$ is a 56-manifold and so also is $\mathbf{C}P^{28}$. The connected sum construction gives us a new 56-manifold $M(2) = 2R(2\phi)\#\mathbf{C}P^{28}$. Starting all over again, take $3\phi : S^{23} \to S^{13}$ and build another 56-manifold, $M(3) = 2R(3\phi)\#\mathbf{C}P^{28}$.

The following statement is very easy to write and very difficult to prove: $M(2)$ and $M(3)$ have the fpp. The proof occupies most of Husseini's paper [**11**] (pp. 922–930). That material is beyond the scope of this paper, but keep in mind that the construction of $M(2)$ and $M(3)$ was governed by the need to be able to compute Lefschetz numbers of maps on them—and to demonstrate that the Lefschetz numbers are always nonzero.

A Fixed Point Free Map. What remains to be done is to show there is a map from $M(2) \times M(3)$ to itself that has no fixed point, so we can conclude that $M(2) \times M(3)$ does not have the fpp even though the factors do have the fpp.

Let's begin with a few general remarks. If X is a space, A is a subset of X, and we have a map taking X to A, then the map may be viewed as a map of X into itself whose fixed points, if any, certainly must lie in A. One way, then, to build a map on X without fixed points is to find a map of A to itself that is without fixed points and extends to a map of X to A (that means there is a map of X to A whose restriction to A is the map we started with). Actually, the map of A to itself can even have fixed points, provided the spaces X and A are well-behaved and the map is homotopic to a map on A without fixed points. The reason is that by a standard result (the Homotopy Extension Theorem, see [**20**]), if the given map extends to a map on X, so does any map homotopic to it, in particular the one without fixed points. For one last generality, suppose $X \subseteq Y$ and there is a retraction $r : Y \to X$. Let $f : X \to A$ be a map extending the map of A to itself without fixed points, then the composition

$$Y \xrightarrow{r} X \xrightarrow{f} A \subseteq X \subseteq Y$$

is a map of Y to itself without fixed points because it could only have them on A, and we know there are none on that subset.

In the present circumstances Y is of course $M(2) \times M(3)$ and I take $X = C(2\phi) \times C(3\phi)$. I claim I have a retraction $r: X \to Y$. Notice that $C(2\phi)$ is a subset of $R(2\phi)$ and therefore of $2R(2\phi)$. When I removed the interior of a copy of I^{56} to form the connected sum $M(2) = 2R(2\phi)\#\mathbf{C}P^{28}$ I could easily avoid $C(2\phi)$. Using the maps I kept track of during the constructions above gives the composition

$$M(2) \xrightarrow{\kappa} 2R(2\phi) \xrightarrow{\pi} R(2\phi) \xrightarrow{\rho} C(2\phi)$$

which is a retraction. Since there is a corresponding retraction of $M(3)$ onto $C(3\phi)$, I've established the claim.

To identify the subset of $C(2\phi) \times C(3\phi)$ that I'll call A, begin by noticing that the spherical complex structure $C(2\phi) = 1 \cup_c I^{13} \cup_{2\phi} I^{24}$ permits us to view $S^{13} = 1 \cup_c I^{13}$ as a subset of $C(2\phi)$. There is a sphere S^{13} in $C(3\phi)$ also, and thus a copy of $S^{13} \times S^{13}$ in $X = C(2\phi) \times C(3\phi)$. Let A be the copy of S^{13} in X that corresponds to the diagonal in $S^{13} \times S^{13}$ (the set of pairs (x, x) for $x \in S^{13}$). The antipodal map $\alpha: S^{13} \to S^{13}$ defined by $\alpha(x) = -x$ is obviously without fixed points. As I suggested above, I won't try to show directly that α extends to a map from X to S^{13}. Instead, I'll construct a map $a: S^{13} \to S^{13}$ which has fixed points but which (1) extends to a map on X and (2) is homotopic to α. Then the Homotopy Extension Theorem implies that α itself does extend and so, by the general remarks, $M(2) \times M(3)$ admits a map without fixed points.

To begin the construction of the map a, recall from Section 3 that a sphere is a suspension:

$$S^n = \Sigma S^{n-1} = \left(S^{n-1} \times [-1, 1]\right)/\sim$$

with nontrivial equivalence classes $-\mathbf{1} = S^{n-1} \times \{-1\}$ and $\mathbf{1} = S^{n-1} \times \{1\}$. So a sphere of any dimension can be obtained from the circle S^1 by repeated suspensions $S^n = \Sigma^{n-1}S^1$ $(= \Sigma\Sigma \cdots \Sigma S^1)$. Now letting $g: S^m \to S^m$ by any map, I can "suspend" it to a map $\Sigma g: S^{m+1} \to S^{m+1}$ by letting $\Sigma g(x, t) = (g(x), t)$ for $-1 < t < 1$ and extending to $-\mathbf{1}$ and $\mathbf{1}$ as in Fig. 11. If we start

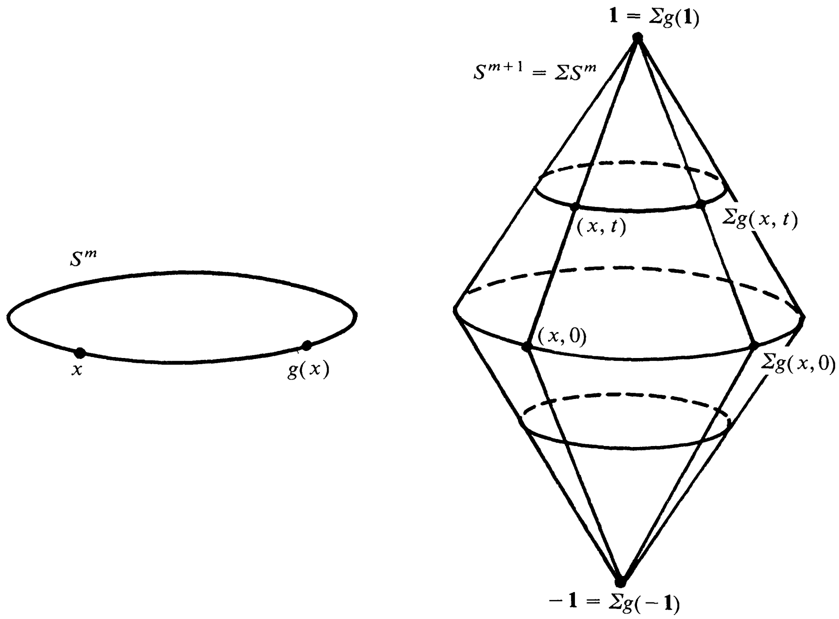


FIG. 11. The suspension of a map.

[Continued on p. 677.]

with a map $g: S^1 \to S^1$ we can produce a map $\Sigma^{n-1}g: S^n \to S^n$ just by iterating this procedure. In particular, think of S^1 as the unit circle in the complex numbers ($S^1 = \{z \in \mathbf{C} \mid \|z\| = 1\}$). For every integer k there is a map $p_k: S^1 \to S^1$ defined by $p_k(z) = z^k$ and thus maps $\Sigma^{n-1}p_k: S^n \to S^n$. The algebraic topology of these maps is well understood.

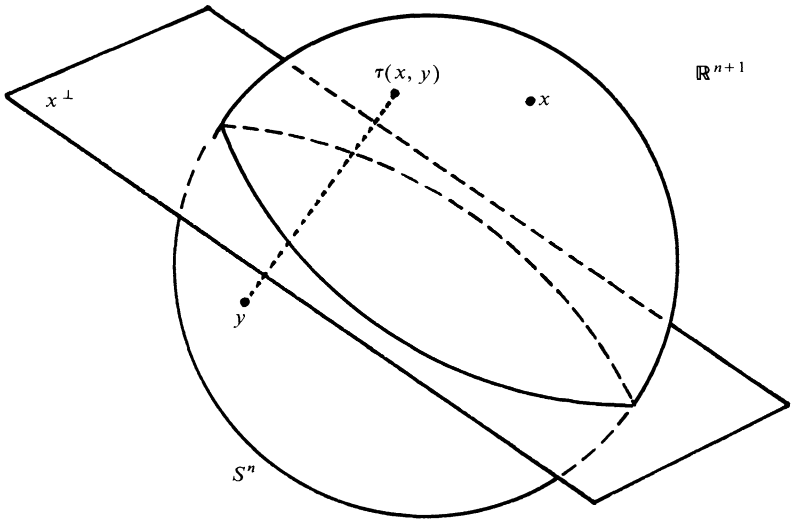


FIG. 12. The map τ.

Next, let $\tau: S^n \times S^n \to S^n$ be the map pictured in Fig. 12, where now we think of S^n as the set of vectors in $\mathbf{R}^{n+1}$ of unit length. The subspace $x^\perp$ consists of all vectors in $\mathbf{R}^{n+1}$ orthogonal to x. Then $\tau(x, y)$ is defined to be the point of S^n obtained by reflecting y through $x^\perp$. To be precise, let $x \cdot y$ denote the usual dot product of vectors, then

$$\tau(x, y) = y - (2x \cdot y)x$$

An important property of the map τ is its lack of symmetry with respect to the two independent variables x and y. To visualize the difference, first let's fix x at some value x_0, then we can think of $\tau(x_0, y)$ as the map from S^n to itself that reflects the entire sphere through the fixed subspace $x_0^\perp$ of vectors orthogonal to x_0. Now let's repeat the experiment, this time holding y fixed at some y_0. The restriction $\tau(x, y_0)$ is more difficult to imagine as a map on S^n. You might think of this map as "scattering" y_0 all over S^n; y_0 is reflected through different subspaces $x^\perp$ which vary with the choice of x. Since x and $-x$ have the same orthogonal subspace, y_0 is sent to every point twice. The lack of symmetry is reflected in the algebraic topology of the map τ (see **[21]**).

Let $\Delta: S^{13} \to S^{13} \times S^{13}$ be the diagonal map $\Delta(x) = (x, x)$ and define a to be the composition

$$S^{13} \xrightarrow{\Delta} S^{13} \times S^{13} \xrightarrow{\Sigma^{12}p_{-3} \times \Sigma^{12}p_2} S^{13} \times S^{13} \xrightarrow{\tau} S^{13}.$$

The map $\Sigma^{12}p_{-3}$ extends to a map of $C(2\phi)$, that is, it extends to the cell I^{24} attached by the map $2\phi: S^{23} \to S^{13}$. The existence of such an extension depends on some algebraic topology but, intuitively, we can notice that $(-3)(2) = -6$ and remember that 6ϕ is homotopic to a constant map. Surely, a constant map can be extended to anything we like. There's somewhat more to the formal argument of course. Similarly, $\Sigma^{12}p_2$ extends to $C(3\phi)$ (notice that $(2)(3) = 6$) so a does extend to a map on $C(2\phi) \times C(3\phi)$.

The final argument, that $a: S^{13} \to S^{13}$ is homotopic to the antipodal map α, is a nice exercise in basic algebraic topology. The lack of symmetry of τ that I was just emphasizing is essential to

the argument. *This* is the point where it is crucial that the two manifolds $M(2)$ and $M(3)$ be different (recall the unsolved problem at the end of Section 2). It is precisely the interplay of properties of 2ϕ, 3ϕ, and τ that makes a homotopic to α. I can't be more specific without an excursion into the algebraic topology (which can be found on page 574 of [**1**]), but it's not hard to show that there is no hope of building a map like a homotopic to α and extending to, for instance, $C(2\phi) \times C(2\phi)$ rather than $C(2\phi) \times C(3\phi)$.

References

1. G. Bredon, Some examples for the fixed point property, Pacific J. Math., 38 (1971) 571–575.

2. R. Brown, The Lefschetz Fixed Point Theorem, Scott, Foresman, Glenview, IL, 1971.

3. E. Connell, Properties of fixed point spaces, Proc. Amer. Math. Soc., 10 (1959) 974–979.

4. E. Dyer, A fixed point theorem, Proc. Amer. Math. Soc., 7 (1956) 662–672.

5. E. Fadell, Recent results in the fixed point theory of continuous maps, Bull. Amer. Math. Soc., 76 (1970) 10–29.

6. ______, Some examples in fixed point theory, Pacific J. Math., 33 (1970) 89–100.

7. ______, Problem, Fixed Point Theory (Proceedings, Sherbrooke 1980), Springer, Berlin, 1981.

8. M. Greenberg, Lectures on Algebraic Topology, Benjamin, New York, 1967.

9. H. Hopf, Vectorfelder in n-dimensionalen Mannigfaltigkeiten, Math. Ann., 96 (1927) 225–250.

10. S. T. Hu, Homology Theory, Holden-Day, San Francisco, CA, 1966.

11. S. Husseini, The product of manifolds with the f.p.p. need not have the f.p.p., Amer. J. Math., 99 (1977) 919–931.

12. R. Jerrard, Homology with multiple-valued functions applied to fixed points, Trans. Amer. Math. Soc., 213 (1975) 407–427.

13. V. Klee, An example related to the fixed-point property, Niew Arch. Wisk., 8 (1960) 81–82.

14. R. Knill, Cones, products and fixed points, Fund. Math., 60 (1967) 35–46.

15. K. Kuratowski, Problem 49, Fund. Math., 15 (1930) 356.

16. W. Lopez, An example in the fixed point theory of polyhedra, Bull. Amer. Math. Soc., 73 (1967) 922–924.

17. J. Munkres, Elementary Differential Topology, Princeton Univ. Press, Princeton, NJ, 1963.

18. C. Rourke and B. Sanderson, Introduction to Piecewise Linear Topology, Springer, Berlin, 1972.

19. R. Sorgenfrey, On the topological product of paracompact spaces, Bull. Amer. Math. Soc., 53 (1947) 631–632.

20. E. Spanier, Algebraic Topology, McGraw-Hill, New York, NY, 1966.

21. N. Steenrod and D. Epstein, Cohomology Operations, Princeton Univ. Press, Princeton, NJ, 1962.

22. H. Toda, Composition Methods in Homotopy Groups of Spheres, Princeton Univ. Press, Princeton, NJ, 1962.

23. J. Vick, Homology Theory, Academic Press, New York, 1973.

24. F. Wecken, Fixpunktklassen III. Mindestzahlen von Fixpunkten, Math. Ann., 118 (1942) 544–577.

25. T. Worosz, Multiple-valued functions and their applications to the fixed point properties of product spaces, Bull. Malaysian Math. Soc., 1 (1978) 101–109.

MISCELLANEA

THE ATOM BOMB GAME

86. Here is a game in the sense of von Neumann that has no "value" in the sense of game theory, but it is thought provoking and deserves to be better known. Each of two players, A and B, picks a number in the unit interval [0, 1]. If the sum of the two numbers is less than or equal to 1, the player with the larger number kills the other; in case of a tie they both survive; if the sum is greater than 1, the referee kills them both.

—The game was known to Irving Kaplansky in the 1950's; he now thinks he first heard it from Paul Erdős.

CENTER SECTION

(Vol. 99, No. 9, November 1982)

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S. Overcoming Math Anxiety. Sheila Tobias. Houghton Mifflin, 1980, 284 pp, $6.95 (P). [ISBN: 0-395-29088-0] Paperback edition of the now-classic brief on the sources of defeat and terror that afflict millions of people. (First Edition, TR, May 1979; ER, February 1980.) LAS

General, S*, L*. Infinity and the Mind: The Science and Philosophy of the Infinite. Rudy Rucker. Birkhauser Boston, 1982, x + 342 pp, $29.95. [ISBN: 3-7643-3034-1] A wide-ranging discourse on the age-old dilemma of the infinite, intertwining traditional mathematical aspects (e.g, Cantor's cardinals, Robinson's infinitesimals, Conway's surreal numbers, Gödel incompleteness theorem) with philosohical roots (e.g., Pythagoreanism, absolute vs. potential, logical paradoxes), and pseudo-science (e.g., mysticism, speculation). The exposition is equally wide-ranging, from popular to technical, mixing imaginative analogies and metaphors with logical notation and symbolism. This unique volume has great potential for interesting non-mathematicians in twentieth century set theory, as well as for misleading naive readers concerning the boundaries between science and speculation, between knowledge and mysticism. LAS

General, S, P**, L**.** The Fractal Geometry of Nature. Benoit B. Mandelbrot. WH Freeman, 1982, 460 pp. [ISBN: 0-7167-1186-9] An extensive revision of Mandelbrot's 1977 Fractals, Form, Chance, and Dimension (TR, August-September 1977), replete with impressive color graphics (e.g., a fractal view of earth from a space mission that never was), and applications to a wide range of scientific applications. Mandelbrot's thesis is that fractals complement calculus as descriptions of nature--the one representing jaggedness, the other smoothness. This volume provides impressive evidence of this view--mathematical foundations, visual representations, extensive references to the scientific literature--as well as fascinating personal asides by the author concerning his experiences and opinions at the center of this exciting new application of mathematics and computer graphics. LAS

Elementary. Instructor's Manual for Elementary School Mathematics and How to Teach It. Eugene D. Nichols, Merlyn J. Behr. Holt, Rinehart & Winston, 1982, iv + 164 pp, (P). [ISBN: 0-03-042006-7] Detailed solutions to selected exercises and chapter tests to support the text Elementary School Mathematics and How to Teach It. JJ

Elementary, T(13: 1). College Algebra with Applications. Mustafa A. Munem, David J. Foulis. Worth Pub, 1982, xi + 410 pp, $18.95. [ISBN: 0-87901-170-X] This appears to be a straightforward and correct treatment of the usual college algebra topics plus introductions to matrix algebra, infinite series, and conic sections. Learning and teaching supplements are available. FLW

Precalculus, T(13: 1). Algebra and Trigonometry with Applications. Mustafa A. Munem, David J. Foulis. Worth Pub, 1982, xii + 589 pp, $19.95. [ISBN: 0-87901-133-5] In addition to the traditional topics and development of skills, the text gives special attention to applications (with index) and models in the life sciences, the social sciences, the physical sciences, economics, and business. A student study guide is available. JJ

Precalculus, T*(13: 1). Trigonometry with Calculators. Marshall D. Hestenes, Richard O. Hill, Jr. Prentice-Hall, 1982, xii + 274 pp, $18.95. [ISBN: 0-13-930859-8] Text incorporates instructions on calculator use but emphasizes understanding of concepts and importance of estimating answers to check work. Covers standard trigonometry topics plus complex numbers and logarithms and exponentials. Extensive exercise sets include many word problems. Adapted from authors' Algebra and Trigonometry with Calculators. KS

Precalculus, T(13-14: 1), S, L. Technical Mathematics Through Applications. Juliana Corn, Tony Behr. Saunders Coll Pub, 1982, xv + 585 pp, $23.95. [ISBN: 0-03-057721-7] Intended for engineering technology students with a year of high school algebra. Topics similar to those of precalculus, with applied focus--e.g., polar coordinates, vectors, graph-sketching on semi-log paper, complex numbers, calculator usage. Most applications are to physics (electricity and kinematics) though no

knowledge of physics is assumed. Four review chapters. Many worked-out "story problems." PZ

Precalculus, T(13: 1). Trigonometry. Leonard I. Holder. Wadsworth Pub, 1982, x + 271 pp, $18.95. [ISBN: 0-534-01014-8] A modest title for a book which has applications to geometry, vectors, velocity, and force. Well-chosen problems, including many for calculator. Including chapters on logarithms, complex numbers, and polar coordinates, it should be considered for any pre-calculus course. RBK

Precalculus, T(13). Algebra & Trigonometry. Thomas W. Hungerford, Richard Mercer. Saunders Coll Pub, 1982, x + 678 pp, $21.95. [ISBN: 0-03-059519-3] A basic algebra and precalculus text. Topics are (in order): algebra, equations and inequalities, analytic geometry, functions, polynomial, rational, exponential, and logrithmic functions, trigonometry, and complex numbers. SG

Education, T(13: 1, 2). Modern Mathematics: An Elementary Approach, Fifth Edition. Ruric E. Wheeler. Brooks/Cole, 1981, xviii + 603 pp, $19.95. [ISBN: 0-8185-0430-7] Text for elementary school teachers and liberal arts majors covering number systems, elementary number theory, geometry, computers, probability and statistics. Changes in this edition include greater emphasis on problem solving and the addition of problems from elementary school texts, biographical sketches, flowcharts for procedures, and special projects. (TR, Second Edition, November 1970; Third Edition, October 1973; Fourth Edition, December 1977.) KS

History, S, P, L. The Land of Stevin and Huygens. Dirk J. Struik. Stud. in History of Modern Sci., V. 7. D Reidel Pub, 1981, xx + 162 pp, $18.50 (P); $34. [ISBN: 90-277-1237-9; 90-277-1236-0] Survey of science and technology (including mathematics) in the Netherlands from about 1580 until 1700. Translation with some revisions of third Dutch edition. KS

History, S?(13-16), P? Vedic Mathematics or Sixteen Simple Mathematical Formulae from the Vedas. Jagadguru Swami Bharati Krishna Tirtha. Motilal Banarsidass, 1978, lviii + 367 pp, Rs. 40 (P). Reprint of a 1965 book. Collection of nonstandard mathematical shortcuts, covering topics from arithmetic to calculus, written by an Indian holy man and genius. Claimed to be obtained by "intuitive revelation" and based on sutras (aphorisms or formulae) from the Vedas (most ancient of Indian scriptures). Reconstruction of parts of a lost sixteen volume work (one for each sutra) supposedly covering all of mathematics. RSK

History, L*. Jordanus de Nemore: De numeris datis. Barnabas Bernard. U of Calif Pr, 1981, xii + 212 pp, $37.50. [ISBN: 0-520-04283-2] The first critical edition (with English translation) of the earliest European treatise on advanced algebra, which was written in the late thirteenth century by "one of the most prestigious natural philosophers of that era. A transitional work between those of al-Khwarizmi and Leonardo da Pisa, De numeris datis is the earliest treatise to make extensive use of letters to represent numbers. The work consists of several score propositions of algebra (e.g., "If the ratio of two numbers together with the sum of their squares is known, then each can be found"), each followed by a general algorithm (using letters) and a specific example. LAS

Foundations, S*(15-17), P, L. Modern Logic--A Survey: Historical, Philosophical, and Mathematical Aspects of Modern Logic and its Applications. Ed: Evandro Agazzi. D Reidel Pub, 1981, viii + 475 pp, $68.50. [ISBN: 90-277-1137-2] Proceedings of a conference on "Modern Logic" held in Rome in September 1976. Papers survey branches of logic, relation of logic to mathematics and other scientific disciplines, and philosophical issues. Level of papers varies, but they are generally less technical than those in Handbook of Mathematical Logic. Bibliography follows each article. KS

Foundations, P. Ontological Economy: Substitutional Quantification and Mathematics. Dale Gottlieb. Oxford U Pr, 1980, vii + 166 pp, $28. [ISBN: 0-19-824420-7] First part presents modification of Quine's criterion of ontological commitment to allow for substitutional quantification and develops method for reducing ontological commitments. Second part applies this method to construct an ontologically neutral language for arithmetic of natural and rational numbers. KS

Foundations, T(13-15: 1), S, L. The Foundations of Analysis: A Straightforward Introduction: Book 1, Logic, Sets and Numbers. K.G. Binmore. Cambridge U Pr, 1980, x + 131 pp, $24.95; $12.95 (P). [ISBN: 0-521-23322-4; 0-521-29915-2] Informal introduction to logical and algebraic foundations of analysis. Covers basic symbolic logic and set theory, countability, and axiomatic and constructive approaches to number systems. Optional sections briefly present more advanced topics including the continuum hypothesis. Standard exercises. KS

Foundations, P. To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism. Ed: J.P. Seldin, J.R. Hindley. Academic Pr, 1980, xxv + 606 pp, $64. [ISBN: 0-12-349050-2] Festschrift in honor of Curry's 80th birthday. Most papers are technical; a few expository articles describe Curry's program. Includes brief biography of Curry and complete bibliography of his work. KS

Foundations, P. Lecture Notes in Mathematics-890: Model Theory and Arithmetic. Ed: C. Berline, K. McAloon, J.-P. Ressayre. Springer-Verlag, 1981, vi + 306 pp, $18 (P). [ISBN: 0-387-11159-X] Since the mid-1970's there has been a burgeoning in the study of axiom systems for arithmetic, particularly by the application of semantic methods (e.g., nonstandard models of arithmetic). This volume of papers grew out of a year-long seminar series at the Université Paris VII, 1979/80. GHM

Graph Theory, P. The Theory and Applications of Graphs. Ed: G. Chartrand, et al. Wiley, 1981, xvi + 611 pp, $29.95. [ISBN: 0-471-08473-5] Proceedings of Fourth International Conference held in May

1980 at Western Michigan University. Topics include tournaments, Ramsey theory, Hamiltonian graphs, colorings, trees and applications to traffic phasing, computer networks, linguistics and chemistry. List of open problems. KS

Algebra, T(13-15: 1, 2), L. Modern Algebra: A Natural Approach, with Applications. C.F. Gardiner. Halsted Pr, 1981, 288 pp, $59.38. [ISBN: 0-470-27115-9] Expensive, readable, unusual text for introductory abstract algebra course. Covers functions and relations, algebra of complex numbers, linear algebra, quadratic forms, Euclidean domains. Includes a few applications to physics and mathematics. Most exercises involve computation or verification; complete solutions at end of book. KS

Algebra, T(15-17: 1), L. Introduction to Group Theory. Hans Schwerdtfeger. Noordhoff Intern, xvi + 238 pp, $25 (P). [ISBN: 90-286-0495-2] Text for introductory course in group theory. Covers basic concepts plus direct products, groups of automorphisms, group extensions, series of subgroups, Sylow theorems, structure of finite abelian groups. Many sections motivated by author asking questions. Exercises and examples intermixed at end of each section; several deal with geometrical groups. KS

Algebra, T*(15-17: 1), S, L*. Introduction to the Theory of Error-Correcting Codes. Vera Pless. Wiley, 1982, xi + 169 pp, $22.95. [ISBN: 0-471-08684-3] An introduction to linear block codes for students in mathematics, computer science and electrical engineering which assumes knowledge of linear algebra. Topics include Hamming codes, cyclic codes, B.C.H. codes, weight distributions, designs, glueing. Develops necessary field theory. Seminar based on this text would be nice sequel to abstract algebra course. KS

Algebra, T(16-17: 1), S, L. A Course in Universal Algebra. Stanley Burris, H.P. Sankappanavar. Grad. Texts in Math., V. 78. Springer-Verlag, 1981, xvi + 276 pp, $29.80. [ISBN: 0-387-90578-2] An interesting introduction to universal algebra which assumes familiarity with basic abstract algebra and point set topology. Covers lattices and Boolean algebras, fundamental concepts of universal algebra, Boolean powers and discriminator varieties, connections with model theory, and applications to Latin squares and finite state acceptors. Contains summary of current research activity and extensive bibliography. KS

Algebra, S(17-18), P. Examples of Commutative Rings. Harry C. Hutchins. Polygonal Pub, 1981, vii + 167 pp, $13.75. [ISBN: 0-936428-05-8] Part I contains summary of definitions and results (no proofs) from commutative ring theory. Part II presents 180 examples of commutative rings and lists their major properties. Assumes basic knowledge of ring theory. Annotated bibliography. KS

Algebra, P. Places and Valuations in Noncommutative Ring Theory. Jan Van Geel. Pure and Appl. Math., V. 71. Dekker, 1981, ix + 112 pp, $19.50 (P). [ISBN: 0-8247-1572-1] Monograph presents generalization of valuation theory for noncommutative rings using notion of a prime pair. Special attention given to central simple algebras and theory of orders. KS

Calculus, T*(13: 1-3). Calculus and Analytic Geometry. C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1982, xvi + 895 pp, $29.95. [ISBN: 0-13-111609-6] Distinctive features include: derivative introduced before formal limits; integral defined using Riemann sums; wide range of interesting applications both in text and exercises; carefully worded definitions and theorems, but more conversational tone elsewhere. Readable, with good problems. JRG

Calculus, T(13: 1). Calculus for Management, Life, and Social Sciences, Second Edition. Raymond A. Barnett. Dellen Pub, 1981, xiii + 385 pp, $20.95. [ISBN: 0-89517-025-6] Effective use of a second color in text and diagrams. Many good applications to business-economics, life sciences, and social sciences. Well annotated steps in examples. Includes multivariable max-min problems. No proofs. (TR, First Edition, June-July 1978.) RBK

Calculus, T(13-14: 3), L. Calculus with Analytic Geometry, Third Edition. Howard E. Campbell, Paul F. Dierker. Prindle, Weber & Schmidt, 1982, xi + 924 pp. [ISBN: 0-87150-331-X] Apart from minor editorial changes, the main revision of the previous editions has come in "geometric motivation," particularly in the multivariable calculus. Also new examples and exercises have been added. (TR, Second Edition, June-July 1978.) JS

Calculus. Mathematik heute: Grundkurs Analysis: Lösungen und didaktisch-methodischer Kommentar. Hermann Athen, Heinz Griesel. Hermann Schroedel, 1981, (P). V. 1, 142 pp [ISBN: 3-507-83182-1]; V. 2, 192 pp. [ISBN: 3-507-83183-X] Problem solutions and teacher's commentary to be used with two volumes of the same name. (Volume 1, TR, April 1979; Volume 2, TR, March 1981.) JD-B

Real Analysis, T(14-15: 1), S, L. The Foundations of Analysis: A Straightforward Introduction, Book 2: Topological Ideas. K.G. Binmore. Cambridge U Pr, 1981, xii + 249 pp, $39.50; $16.95 (P). [ISBN: 0-521-23350-X; 0-521-29930-6] This text is designed to bridge the gap between elementary analysis courses and more advanced courses in analysis. It does indeed present the topological foundations of analysis in a well organized, well thought out manner. The style is crisp but easy to read and the text includes ample exercise sets. The hardcover cost is excessive. PH

Real Analysis, T(14-15), S, L. Elementary Mathematical Analysis, Second Edition. Colin W. Clark. Wadsworth Pub, 1982, xi + 259 pp, $19.95. [ISBN: 0-534-98018-X] A second edition of the book The Theoretical Side of Calculus (1972; TR, April 1973; ER, August-September 1974). This edition follows the first quite closely, but contains some additional material on series, additional exercises, and an appended application to the theory of differential equations. PH

Real Analysis, T(16-17: 1, 2), S. L. Principles of Real Analysis. Charalambos D. Aliprantis, Owen Burkinshaw. Elsevier North Holland, 1981, xii + 285 pp, $27.95. [ISBN: 0-444-00448-3] A well-written introduction to measure and integration with ample exercises and examples. The integral is introduced via an interesting adaptation of the method of Daniell: a function f is Lebesgue integrable means that f = g-h where each of g and h is the increasing limit of step functions. However, a fairly general discussion of measure precedes this formulation and this allows the authors to develop the theory in a fair degree of generality. The text emphasizes and uses the algebraic structure of the various function spaces considered and this certainly enhances the appeal. The book is expensive, but the format, paper, and printing are first class. PH

Real Analysis, T(18: 2), P. Singular Integral Equations: Boundary Problems of Function Theory and Their Application to Mathematical Physics. N.I. Muskhelishvili. Trans: J.R.M. Radok. Noordhoff Intern, 1977, 447 pp, $50. [ISBN: 90-01-60700-4] A revised translation from the Russian text and a reprint of the 1958 edition. This monograph is an edited systematic study of that portion of Cauchy type integrals and singular integrals investigated by the author and his students. Applications to numerous problems of potential theory, elasticity and other areas of mathematical physics form an appreciable portion of the text. PH

Differential Equations, P. Abstract Differential Equations. S.D. Zaidman. Research Notes in Math., V. 36. Fearon Pitman Pub, 1979, 130 pp, $15.95 (P). [ISBN: 0-8224-8427-7] A study of linear differential equations in Banach and Hilbert spaces in which coefficients are linear unbounded operators. JG

Numerical Analysis, T(18: 1), P. Lectures on Numerical Methods for Non-Linear Variational Problems. R. Glowinski. Springer-Verlag, 1980, vii + 240 pp, $7.70 (P). [ISBN: 0-387-08774-5] Lecture notes on parabolic and elliptic variational inequalities; applied to non-linear variational equations; finite element, augmented Lagrangian and general iterative methods; problems in mechanics and fluid dynamics. RWN

Numerical Analysis, T(14-15). Numerical Methods for Engineering Application. Joel H. Ferziger. Wiley, 1981, xii + 270 pp, $25.95. [ISBN: 0-471-06336-3] Intended for engineering students. A compendium of popular methods for interpolation, integration, ordinary differential equations, and (about half of the book) partial differential equations. Some practical advice; no analysis. RWN

Numerical Analysis, P. Numerical Analysis of Variational Inequalities. Roland Glowinski, Jacques-Louis Lions, Raymond Trémolières. Stud. in Math. and its Appl., V. 8. Elsevier North Holland, 1981, xxix + 776 pp, $109.75. [ISBN: 0-444-86199-8] Considers numerical solutions to partial differential equations or variational problems in which the solution unit satisfies inequalities. Classifies problems; develops, analyzes, and applies various methods. Significant examples. Over 200 pages on progress since the 1974 French volumes. RWN

Numerical Analysis, P. The Application and Numerical Solution of Integral Equations. Ed: Robert S. Anderssen, Frank R. de Hoog, Mark A. Lukas. Mechanics: Analysis, V. 6. Sijthoff & Noordhoff, 1980, x + 259 pp, $27.50. [ISBN: 90-286-0450-2] 12 review papers on the classification of integral equations, general numerical methods for each class, and applications. RWN

Numerical Analysis, S(17-18), P. Padé Approximants. George A. Baker, Jr., Peter Graves-Morris. Addison-Wesley, 1981. Part I: Basic Theory, xx + 325 pp, $32.50 [ISBN: 0-201-13512-4]; Part II: Extensions and Applications, xviii + 215 pp, $29.50. [ISBN: 0-201-13513-2] Part I contains definitions, methods of calculation, applications to numerical methods, convergence theory and the relationship to continued fractions and Stieltjes series. Part II contains generalizations and applications to integral equations, scattering theory, quantum field theory as well as numerical analysis. Intended for a general mathematical audience. RWN

Functional Analysis, T(18: 1), S, P. Lecture Notes in Mathematics-853: Order and Convexity in Potential Theory: H-Cones. Nicu Boboc, Gheorghe Bucur, Aurel Cornea. Springer-Verlag, 1981, iv + 286 pp, $16.80 (P). [ISBN: 0-387-10692-8] Much of the information concerning potential functions can be derived from the algebraic and order-theoretic properties. This monograph studies potential theory from the more general setting obtained by abstracting these properties from the set of potentials. The four algebraic properties of potentials fundamental to this study are articulated in the fairly elementary definition of H-cone. The investigation of H-cones in the context of more familiar mathematical constructs leads to an interesting perspective of some potential-theoretic results. PH

Functional Analysis, S(17-18), P. Nuclear and Conuclear Spaces: Introductory Course on Nuclear and Conuclear Spaces in the Light of the Duality ′Topology-Bornology′. Henri Hogbe-Nlend, Vincenzo Bruno Moscatelli. Math. Stud., V. 52. Elsevier North-Holland, 1981, x + 275 pp, $41.50 (P). [ISBN: 0-444-86207-2] An introduction to the theory of nuclear and conuclear spaces. Includes a study of the duality between the two "in the light of the duality ′topology-bornology′." Sets of exercises are included with each chapter. Comes with the usual North-Holland price tag. PH

Analysis, P. Random Fourier Series with Applications to Harmonic Analysis. Michael B. Marcus, Gilles Pisier. Annals of Math. Stud., No. 101. Princeton U Pr, 1981, 150 pp, $17.50; $7 (P). Author′s abstract: "Necessary and sufficient conditions are obtained for the a.s. uniform convergence of random Fourier series on locally compact Abelian groups and on compact non Abelian groups. Many related results such as the central limit theorem are obtained. The methods developed are used

to study questions in harmonic analysis which are not intrinsically random." PH

Analysis, T(15). Multivariate Calculus. Lawrence J. Corwin, Robert H. Szczarba. Pure and Appl. Math., No. 64. Dekker, 1982, xi + 524 pp, $49.50. [ISBN: 0-8247-6962-7] Making use of linear algebra as prelude to defining the derivative as a linear transformation, but using it only to the extent that it actually clarifies an approach that is clearly a classical, rigorous approach to what used to be called advanced calculus. Attractive! AWR

Algebraic Geometry, P. Lectures on Hilbert Modular Surfaces. Friedrich Hirzebruch, Gerard van der Geer. Pr U Montreal, 1981, 193 pp, $14 (P). [ISBN: 2-7606-0562-0] Lectures by Hirzebruch on algebraic surfaces and by van der Geer on Hilbert modular surfaces. Many interesting examples and an extensive bibliography. SG

Algebraic Geometry, P. Géométrie Analytique Rigide et Applications. Jean Fresnel, Marius van der Put. Birkhauser Boston, 1981, xii + 215 pp, $10. [ISBN: 3-7643-3069-4] An introduction to rigid analytic spaces together with applications to elliptic curves and projective varieties. SG

Algebraic Topology, P. K-Theory of Forms. A. Bak. Annals of Math. Stud., No. 98. Princeton U Pr, 1981, viii + 268 pp, $20; $8.50 (P). Covers quadratic, hermitian and even hermitian forms and their algebraic K-theories. Unifies theories by introducing a form parameter which is a particular additive subgroup of a ring with involution. KS

Algebraic Topology, T(15: 1), S, L. Graphs, Surfaces and Homology: An Introduction to Algebraic Topology, Second Edition. P.J. Giblin. Chapman & Hall, 1981, xvii + 329 pp, $15.95 (P). [ISBN: 0-412-23900-0] Only minor changes from the First Edition (TR, June-July 1977), including remarks about four-color theorem. JG

Topology, P. Topology and Order Structures, Part 1. Ed: H.R. Bennett, D.J. Lutzer. Math. Centre Tracts, V. 142. Math Centrum, 1981, iii + 184 pp, Dfl. 24,15 (P). [ISBN: 90-6196-228-5] In August, 1980 NATO and Texas Tech University sponsored a workshop on topology and linear order for specialists in this area. The present volume represents the contributions of most of the participants together with a few papers by others unable to attend. A subsequent volume is in process. PH

Topology, P. The Closed Graph and P-Closed Graph Properties in General Topology. T.R. Hamlett, L.L. Herrington. Contemporary Math., V. 3. AMS, 1981, xi + 68 pp, $6.80 (P). [ISBN: 0-8218-5004-0] Monograph presents results on conditions under which the graph of a function is closed and the relation of the closed graph property to continuity, minimal topological spaces, compactness, and non-continuous functions. Analogous results are obtained for functions with theta-closed graphs. KS

Operations Research, T(14-15: 1, 2). Management Science: An Introduction to Quantitative Analysis for Management. William E. Pinney, Donald B. McWilliams. Harper & Row, 1982, xxi + 568 pp, $26.50. [ISBN: 0-06-045222-6] Survey of techniques, with emphasis on application and problem recognition. Many examples. Includes review chapters on algebra, matrix algebra, and differential calculus. JRG

Operations Research, T(16-17: 1), L.** Dynamic Programming: Models and Applications. Eric V. Denardo. Prentice-Hall, 1982, xii + 227 pp, $26.95. [ISBN: 0-13-221507-1] Up-to-date account of key ideas of dynamic programming and their uses in operations research. Topics include the shortest route problem, resource allocation, production and inventory control. Final chapter introduces sequential decision processes lacking fixed planning horizons. Supplements on data structures and convexity. Nice choice of end-of-chapter problems, as well as exercises throughout the text. JRG

Operations Research, T(16-17), P, L. Sequencing and Scheduling: An Introduction to the Mathematics of the Job-Shop. Simon French. Math. and its Appl. Ellis Horwood Pub, 1982, x + 245 pp, $49.95. [ISBN: 0-85312-299-7] Through the practical problem of scheduling, the reader is led into much of combinatorial optimization: dynamic programming, integer programming, NP-completeness, and heuristics. Written for the mathematically oriented non-specialist, but likely to be a useful reference for the professional in operations research. AWR

Operations Research, P. Mathematical Programming with Data Perturbations I. Ed: Anthony V. Fiacco. Lect. Notes in Pure & Appl. Math., V. 73. Dekker, 1982, x + 237 pp, $34.50 (P). [ISBN: 0-8247-1543-8] Twelve papers on basic results and applications of sensitivity and stability methods for linear, nonlinear, integer, and stochastic programming. RWN

Optimization, T(16-17: 1). Practical Methods of Optimization: Volume 2, Constrained Optimization. R. Fletcher. Wiley, 1981, ix + 224 pp, $31.95. [ISBN: 0-471-27828-9] Topics include linear, nonlinear, and quadratic programming, as well as a chapter on non-differentiable optimization. Emphasis on practicability. JRG

Optimization, P. Stochastic Dynamic Programming: Successive Approximations and Nearly Optimal Strategies for Markov Decision Processes and Markov Games. J. van der Wal. Math. Centre Tracts, V. 139. Math Centrum, 1981, iii + 251 pp, Dfl. 32,55 (P). [ISBN: 90-6196-218-8] Dynamic programming methods for approximation of the value and determination of nearly-optimal stationary strategies in total-reward and average-reward Markov decision processes and two-person zero-sum Markov games. JRG

Statistics, P. Analysis of Categorical Data: Dual Scaling and its Applications. Shizuhiko Nishisato. Math. Expos., No. 24. U of Toronto Pr, 1980, xiii + 276 pp, $27.50. [ISBN: 0-8020-

5489-7] Presents theory of dual (or optimal) scaling and describes specific applications of the technique to different forms of categorical data, including contingency and response-frequency tables, response-pattern tables, rank-order/paired comparison tables, and multidimensional tables. Includes many numerical examples and a history of this relatively new technique. RSK

Statistics, P*. Handbook of the Normal Distribution. Jagdish K. Patel, Campbell B. Read. Statistics, V. 40. Dekker, 1982, ix + 337 pp, $35. [ISBN: 0-8247-1541-1] Collection of well-documented results (no proofs) relating to the normal distribution, particularly distributional properties. Also includes a brief historical background. Good set of references. RSK

Statistics, T(13: 2). Statistical Techniques in Business and Economics, Fifth Edition. Robert D. Mason. Richard D. Irwin, 1982, xxii + 679 pp, $23.95. [ISBN: 0-256-02656-1] Fifth Edition incorporates self-review examples and chapter exams, some examples using SPSS and MINITAB, and very useful chapter summaries and outlines. It also contains methods for smoothing time series, and normal approximation to the binomial. A big book. (TR, Fourth Edition, November 1978.) RBK

Statistics, P. Einige neuere statistische Verfahren zur Erfassung kausaler Beziehungen zwischen Zeitreihen: Darstellung und Kritik. Gebhard Kirchgässner. Vandenhoeck & Ruprecht, 1981, 190 pp, DM 49 (P). [ISBN: 3-535-11251-3] An account and critique, for specialists, of statistical tests for (Wiener-Granger) causality among time series. JD-B

Statistics, T?(15-16: 1), S, L? Engineering Statistics with a Programmable Calculator. William Volk. McGraw-Hill, 1982, iv + 362 pp, $19.95. [ISBN: 0-07-067552-X] Emphasizes "solutions of statistical problems by means of small, programmable calculators." Many programs supplied for HP and TI calculators. Seems not to cover many basic ideas such as randomness and the central limit theorem. Confuses discrete and continuous situations. FLW

Computer Programming. Programming Basic with the TI Home Computer. Herbert D. Peckham. McGraw-Hill, 1979, xiv + 306 pp, $10.95 (P). [ISBN: 0-07-049156-9] A very gentle introduction to Basic, adapted to the TI home computer. Does not attempt to teach good programming: "The only question is whether your program works or not." Sorting program is faulty. Barely discusses special features of the TI computer. RBK

Computer Programming. Problem-Solving on the TRS-80 Pocket Computer. Don Inman, Jim Conlan. Wiley, 1982, 255 pp, $8.95 (P). [ISBN: 0-471-09270-3] For owners of the TRS-80 pocket computer. Detailed, useful information. Emphasis on using the pocket calculator as a tool in problem solving. RBK

Data Structures, T*(13-18: 1), S. An Introduction to Computer Programming and Data Structures Using MACRO-11. Harry R. Lewis. Reston Pub, 1981, xii + 241 pp, $15.95. [ISBN: 0-8359-3143-9] MACRO-11 is used throughout. General programming concepts are presented and illustrated in a machine level programming context. Standard data structures, parsing, and compiling comprise the second part. Part 3 presents six lengthy programming projects along with solution outlines. Many exercises. Several appendices. Index. RJA

Computer Science, T(15-17: 1), L. Jewels of Formal Language Theory. Arto Salomaa. Computer Sci Pr, 1981, ix + 144 pp, $24.95. [ISBN: 0-914894-69-2] Self-contained, readable introduction to formal language theory focusing on concept of a morphism on a free monoid. Covers repetitions of words, regular languages, codes, decidability, morphic representations. Mentions open problems. Exercises at end of chapters; most involve proofs. KS

Control Theory, P. Dynamic Feedback in Finite- and Infinite-Dimensional Linear Systems. J.M. Schumacher. Math. Centre Tracts, V. 143. Math Centrum, 1981, 175 pp, Dfl. 22,05 (P). [ISBN: 90-6196-229-3] A study of one aspect of mathematical control theory in which change from normal behavior is "eliminated by an integration of the error, which means that the controller brings its own dynamics into the feedback loop." Specifically a direct study of dynamic feedback which uses a "compensator couple," a pair of subspaces one of which is controlled invariant and the other conditional invariant. This general setting generalizes traditional state-space techniques of compensator design and studies when a separation between feedback and observer action is possible. A number of specific applications are investigated. PH

Applications, S(15-16), L. Thinking With Models: Mathematical Models in the Physical, Biological, and Social Sciences. Thomas L. Saaty, Joyce M. Alexander. Pergamon Pr, 1981, xi + 181 pp, $15 (P). [ISBN: 0-08-026474-3] A collection of examples of mathematical models. Includes problems and a chapter on methodology. RWN

Applications (Artificial Intelligence), P. The Process of Question Answering: A Computer Simulation of Cognition. Wendy G. Lehnert. Lawrence Erlbaum Assoc, 1978, ix + 278 pp, $19.95. [ISBN: 0-470-26485-3] Description of a computational model of question answering based on concepts from natural-language processing. Contains summary of necessary background material and comparison with other models. KS

Applications (Artificial Intelligence), P. A Framework for Distributed Problem Solving. Reid G. Smith. Comp. Sci. Artificial Intelligence, No. 10. UMI Research Pr, 1981, xv + 174 pp, $39.95. [ISBN: 0-8357-1218-4] Describes tools for solving ill-structured problems on a distributed processor system. Also discusses goals, the general framework, and other systems. RWN

Applications (Economics), P. Mathematical Modelling of Energy Systems. Ed: Ibrahim Kavrakoglu. Sijthoff & Noordhoff, 1981, xiii + 476 pp. [ISBN: 90-286-0690-4] Typeset papers from the NATO Advanced Study Institute held in June, 1979 at Istanbul, Turkey. Twenty-one articles outline attempts to model energy-economic relations, national energy systems, the world oil market, and electrical systems. AWR

Applications (Engineering), P. Continuum Models of Discrete Systems 4. Ed: O. Brulin, R.K.T. Hsieh. North-Holland, 1981, xvi + 517 pp, $63.75. [ISBN: 0-444-86309-5] Full text of 54 invited contributions to an International Conference held in Stockholm in July 1981, the fourth in a biennial series begun in 1975. Major focus: to penetrate the microscopic world of matter by means of continuum theories. LAS

Applications (Management), T, S, L. Case Studies in Mathematical Modelling. Ed: D.J.G. James, J.J. McDonald. Halsted Pr, 1981, 214 pp, $17.95 (P). [ISBN: 0-470-27177-9] Finding similarly titled books to be vehicles for introducing new mathematics in a form attractive to readers with an attraction to the practical, the studies here make deliberate effort to introduce no new mathematical concepts to an upper division math major, but to focus on the techniques of how to use the tools of math in applied settings. AWR

Applications (Management), T(17: 1), L. Competitive Strategies: An Advanced Textbook in Game Theory for Business Students. Jean-Pierre Ponssard. Elsevier North-Holland, 1981, ix + 211 pp, $42.50. [ISBN: 0-444-86230-7] Introduction to game theory as a tool for analyzing conflicts of interest concerning management problems. A number of case studies; no exercises. JRG

Applications (Physics), P. Mathematical Physics Reviews. Ed: S.P. Novikov. Soviet Sci Reviews. Volume 1 (1980), xi + 207 pp, $49 [ISBN: 3-7186-0019-6]; Volume 2 (1981), x + 269 pp, $70.50. [ISBN: 3-7186-0069-2] These two volumes, the first in a series, contain surveys of recent work by Soviet researchers in rigorous statistical mechanics, the exact solution of certain nonlinear equations, quantum fluctuations of instantons, and stochasticity in nonlinear systems. AO

Applications (Physics), T(18), S, P, L*. The Logic of Quantum Mechanics. Enrico G. Beltrametti, Gianni Cassinelli. Ency. of Math. & its Appl., V. 15. Addison-Wesley, 1981, xxvi + 305 pp, $31.50. [ISBN: 0-201-13514-0] A concise exposition of both classical and contemporary research in the mathematical foundations of quantum mechanics, in three parts: first, a modern exposition of the formalism of quantum mechanics as linear operators on Hilbert space; second, an explication of the constituent mathematical structures and assumptions--the logic of quantum mechanics, focusing on ties to direct empirical evidence; finally, a reconstruction of the Hilbert space formalism from these logical constructs. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

Intermountain Section

The spring meeting of the Intermountain Section was held at the Southern Utah State College in Cedar City on April 16-17, 1982.

Invited Addresses:

"The Way It Was," by Ivan Niven, University of Oregon.
"Paradoxical Coverings of the Real Line," by Ivan Niven, University of Oregon.

Papers and Presentations:

"Some Results and Problems in Intersection Graphs," by L. Kirk Tolman, Brigham Young University.
"Representations of Elements in Distributive Lattices of Finite Breadth," by Henry F. Bunch, Hill Air Force Base.
* "Maximization of Functions of Discrete Parameters," by J. Earl Faulkner, Brigham Young University.

* "Ellipses from a Circular and Spherical Point of View," by Alden Partridge, Ricks College.
"The Barrier Problem," by R. Deane Branstetter, Brigham Young University.
"But I Got the Right Answer," by Paul Yearout, Brigham Young University.
* "Morphisms with Drazin Inverse of Index at Most One," by Donald W. Robinson, Brigham Young University.
* "Surfaces with a Continuous Family of Tangent Planes," by C.E. Burgess, University of Utah. (Will be published as part of a collection on The Scottish Book from Birkhauser Press.)
* "A New Algorithm for Powers of Polynomials by Using a Linear Recurrence Relation," by Donald R. Snow, Brigham Young University. (Available as a copy of an Apple disk.)
"The UMAP Project," by Linda Hill, Idaho State University.

Panel Discussion:

"What Geometry Should be Taught in High School." Moderator: Steven Heath, Southern Utah State College. Panel: Don Tucker, University of Utah; James Case, University of Utah; and Lynn Garner, Brigham Young University.

The M.A.A. video film series, "Mathematics at Work in Society," was introduced by C.E. Burgess.

Northeastern Section

The summer meeting of the Northeastern Section was held at the University of Southern Maine at Gorham, on June 18-19, 1982. Approximately 40 people were in attendance.

Invited Lectures:

"Mathematical and Science Education in the People's Republic of China," by John C. Howe, Rice Memorial High School, Burlington, Vermont.
"Color Symmetry," by Marjorie Senechal, Smith College.
"Three Interesting Problems from Ancient Egyptian Mathematics," by Howard W. Eves, Professor Emeritus, University of Maine.
"Comets, Periodic Orbits, and the Restricted 3-Body Problem," by Edward Belbruno, Boston University.

Metropolitan New York Section

The forty-first annual meeting of the Metropolitan New York Section was held at the Courant Institute of Mathematical Sciences, New York University on Sunday, May 2, 1982, with approximately 200 persons in attendance.

Invited Lectures:

"The Loss of Certainty," by Morris Kline, Professor Emeritus, Courant Institute of Mathematical Sciences, New York University.
"The Fourth Dimension and Computer Animated Geometry," by Thomas J. Banchoff, Brown University.

Short Presentations:

"A Survey of Applications of Stochastic Approximation," by Marvin Yablon, John Jay College of Criminal Justice.
"Planetary Motion Under the Principle of Equivalence," by David Shepulsky, City College of New York.
"Ampere's Remainder for Taylor's Theorem," by A. Novikoff and H. Walker, New York University.
"The Alabama Paradox," by Leonard Nissim, Fordham University.
"A Nonbivalent Borelian Nullset of Power c," by Jay Schiffman, Kean College of New Jersey.
"CAI Project at NYC Tech," by Eli Stern and Henry Africk, New York City Technical College.
"An Improved Lower Bound on the Greatest Element of a Sum," by Noam Elkies, Stuyvesant High School.
"Perfect and Beyond," by Saechin Kim, Bronx High School of Science.
"An Extension of a Theorem of Ramanujan Revisited," by Joseph Arkin, New York Academy of Science.
"On a Class of Path-Independent Integrals in Various Physical Phenomena," by Alan Hoenig, John Jay College of Criminal Justice.
"Tomographic Reconstruction of Convex Sets," by Andrew Markoe, Rider College, and James V. Peters, C.W. Post.
"Newton, Halley, and Who," by Ellis Von Eschen and S.P. Gordon, Suffolk County Community College.
"On Nonmeasurable Sets," by Maurice Machover, St. Johns University.

Awards:

Noam Elkies of Stuyvesant High School received the Charles Salkind Award for the highest regional score in the M.A.A. High School Mathematics competition. Sergey Troyanovsky of SUNY at Stony Brook received the Section Award for the highest regional score in the William Putnam Mathematics Competition.

the argument. *This* is the point where it is crucial that the two manifolds $M(2)$ and $M(3)$ be different (recall the unsolved problem at the end of Section 2). It is precisely the interplay of properties of 2ϕ, 3ϕ, and τ that makes a homotopic to α. I can't be more specific without an excursion into the algebraic topology (which can be found on page 574 of [**1**]), but it's not hard to show that there is no hope of building a map like a homotopic to α and extending to, for instance, $C(2\phi) \times C(2\phi)$ rather than $C(2\phi) \times C(3\phi)$.

References

1. G. Bredon, Some examples for the fixed point property, Pacific J. Math., 38 (1971) 571–575.

2. R. Brown, The Lefschetz Fixed Point Theorem, Scott, Foresman, Glenview, IL, 1971.

3. E. Connell, Properties of fixed point spaces, Proc. Amer. Math. Soc., 10 (1959) 974–979.

4. E. Dyer, A fixed point theorem, Proc. Amer. Math. Soc., 7 (1956) 662–672.

5. E. Fadell, Recent results in the fixed point theory of continuous maps, Bull. Amer. Math. Soc., 76 (1970) 10–29.

6. ______, Some examples in fixed point theory, Pacific J. Math., 33 (1970) 89–100.

7. ______, Problem, Fixed Point Theory (Proceedings, Sherbrooke 1980), Springer, Berlin, 1981.

8. M. Greenberg, Lectures on Algebraic Topology, Benjamin, New York, 1967.

9. H. Hopf, Vectorfelder in n-dimensionalen Mannigfaltigkeiten, Math. Ann., 96 (1927) 225–250.

10. S. T. Hu, Homology Theory, Holden-Day, San Francisco, CA, 1966.

11. S. Husseini, The product of manifolds with the f.p.p. need not have the f.p.p., Amer. J. Math., 99 (1977) 919–931.

12. R. Jerrard, Homology with multiple-valued functions applied to fixed points, Trans. Amer. Math. Soc., 213 (1975) 407–427.

13. V. Klee, An example related to the fixed-point property, Niew Arch. Wisk., 8 (1960) 81–82.

14. R. Knill, Cones, products and fixed points, Fund. Math., 60 (1967) 35–46.

15. K. Kuratowski, Problem 49, Fund. Math., 15 (1930) 356.

16. W. Lopez, An example in the fixed point theory of polyhedra, Bull. Amer. Math. Soc., 73 (1967) 922–924.

17. J. Munkres, Elementary Differential Topology, Princeton Univ. Press, Princeton, NJ, 1963.

18. C. Rourke and B. Sanderson, Introduction to Piecewise Linear Topology, Springer, Berlin, 1972.

19. R. Sorgenfrey, On the topological product of paracompact spaces, Bull. Amer. Math. Soc., 53 (1947) 631–632.

20. E. Spanier, Algebraic Topology, McGraw-Hill, New York, NY, 1966.

21. N. Steenrod and D. Epstein, Cohomology Operations, Princeton Univ. Press, Princeton, NJ, 1962.

22. H. Toda, Composition Methods in Homotopy Groups of Spheres, Princeton Univ. Press, Princeton, NJ, 1962.

23. J. Vick, Homology Theory, Academic Press, New York, 1973.

24. F. Wecken, Fixpunktklassen III. Mindestzahlen von Fixpunkten, Math. Ann., 118 (1942) 544–577.

25. T. Worosz, Multiple-valued functions and their applications to the fixed point properties of product spaces, Bull. Malaysian Math. Soc., 1 (1978) 101–109.

MISCELLANEA

THE ATOM BOMB GAME

86. Here is a game in the sense of von Neumann that has no "value" in the sense of game theory, but it is thought provoking and deserves to be better known. Each of two players, A and B, picks a number in the unit interval [0, 1]. If the sum of the two numbers is less than or equal to 1, the player with the larger number kills the other; in case of a tie they both survive; if the sum is greater than 1, the referee kills them both.

—The game was known to Irving Kaplansky in the 1950's; he now thinks he first heard it from Paul Erdős.

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

L. F. KLOSINSKI
Department of Mathematics, University of Santa Clara, CA 95053

G. L. ALEXANDERSON
Department of Mathematics, University of Santa Clara, CA 95053

A. P. HILLMAN
Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131

The following results of the forty-second William Lowell Putnam Mathematical Competition, held on December 5, 1981, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, five thousand dollars, was awarded to the Department of Mathematics of **Washington University**, St. Louis, Missouri. The members of its winning team were: Kevin P. Keating, Edward A. Shpiz, and Richard A. Stong; each was awarded a prize of two hundred fifty dollars.

The second prize, two thousand five hundred dollars, was awarded to the Department of Mathematics of **Princeton University**, Princeton, New Jersey. The members of its team were: Gregg N. Patruno, David P. Roberts, and Charles H. Walter; each was awarded a prize of two hundred dollars.

The third prize, one thousand five hundred dollars, was awarded to the Department of Mathematics of **Harvard University**, Cambridge, Massachusetts. The members of its team were: Michael J. Larsen, Laurence E. Penn, and Michael Raship; each was awarded a prize of one hundred fifty dollars.

The fourth prize, one thousand dollars, was awarded to the Department of Mathematics of **Stanford University**, Stanford, California. The members of its team were: Richard J. Beigel, Thomas C. Hales, and Kenneth O. Olum; each was awarded a prize of one hundred dollars.

The fifth prize, five hundred dollars, was awarded to the Department of Mathematics of the **University of Maryland**, College Park, Maryland. The members of its team were Ravi B. Boppana, Andrew E. Gelman, and Brian R. Hunt; each was awarded a prize of fifty dollars.

The five highest-ranking individual contestants, in alphabetical order, were **David W. Ash**, University of Waterloo; **Scott R. Fluhrer**, Case Western Reserve University; **Michael J. Larsen**, Harvard University; **Robin A. Pemantle**, University of California, Berkeley; and **Adam Stephanides**, University of Chicago. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of five hundred dollars by the Putnam Prize Fund.

The next five highest-ranking individuals, in alphabetical order, were *Benji N. Fisher*, Harvard University; *Brian R. Hunt*, University of Maryland, College Park; *Kevin P. Keating*, Washington University, St. Louis; *Kenneth O. Olum*, Stanford University; and *Richard A. Stong*, Washington University, St. Louis. Each of these students was awarded a prize of two hundred fifty dollars.

The following teams, named in alphabetical order, received honorable mention: *California Institute of Technology*, with team members Lance J. Dixon, Scott R. Johnson, Zinovy B. Reichstein; *Case Western Reserve University*, with team members Scott R. Fluhrer, Barak A. Pearlmutter, John D. Yeager; *University of Toronto*, with team members David M. Atwood, Ivo Klemes, Philip A. Mansfield; *University of Waterloo*, with team members David W. Ash, A. Michael O'Brien, Gordon J. Sinnamon; and *Yale University*, with team members Alan S. Edelman, Paul N. Feldman, Ronald H. Weinstock.

Honorable mention was achieved by the following thirty-four individuals, named in alphabetical order: *Troy W. Barbee*, Princeton University; *Richard J. Beigel*, Stanford University; *Gary M. Bernstein*, Princeton University; *Christopher J. Bishop*, Michigan State University; *Ravi B. Boppana*, University of Maryland, College Park; *Eric D. Carlson*, Michigan State University; *Stephen J. Curran*, Beloit College; *David L. Des Jardins*, Massachusetts Institute of Technology; *Robert A. Ewan*, Queen's University, Kingston; *Michael V. Finn*, Harvard University; *Nathaniel E. Glasser*, Yale University; *Lin Goldstein*, University of California, Berkeley; *Fred W. Helenius*, Massachusetts Institute of Technology; *Robert J. Holt*, Stanford University; *Irwin L. Jungreis*, Cornell University; *Joe J. Kilian*, Massachusetts Institute of Technology; *Ivo Klemes*, University of Toronto; *Mark K. Maginity*, The Johns Hopkins University; *Evan W. Morton*, Harvard University; *Gregg N. Patruno*, Princeton University; *Mark G. Pleszkoch*, University of Virginia; *Michael Raship*, Harvard University; *David P. Roberts*, Princeton University; *James R. Roche*, University of Notre Dame; *Daniel J. Scales*, Princeton University; *Brian F. Sheppard*, Harvard University; *Edward A. Shpiz*, Washington University, St. Louis; *Carlos T. Simpson*, Harvard University; *Bruce K. Smith*, Princeton University; *John M. Sullivan*, Harvard University; *Pierre Tremblay*, Université Laval; *Jerome V. Walsh*, University of Illinois, Urbana-Champaign; *Charles H. Walter*, Princeton University; *David I. Wolland*, Harvard University.

The other individuals who achieved ranks among the top 100, in alphabetical order of their schools, were: University of Alabama, *Richard B. Borie*; University of Alberta, *Michael P. Lamoureux*; Amherst College, *F. Miller Maley*; California Institute of Technology, *Bradley W. Brock*, *Lance J. Dixon*, *Scott R. Johnson*, *Forrest C. Quinn*, *Zinovy B. Reichstein*; University of California, Los Angeles, *Paul T. Lockhart*; University of California, Santa Barbara, *Eric S. Williams*; Carleton College, *Jon L. Shreve*; Case Western Reserve University, *Barak A. Pearlmutter*, *Susan G. Staples*; University of Chicago, *Daniel J. Goldstein*; Colorado State University, *Tom S. Watts*; Harvard University, *Theodore M. Alper*, *William I. Chang*, *Zachary M. Franco*, *David J. Montana*, *Laurence E. Penn*, *James G. Propp*, *Alfred D. Shapere*, *Gregory B. Sorkin*; Harvey Mudd College, *Matthew G. Hudelson*; Université Laval, *Frederic M. Gourdeau*; University of Maryland, College Park, *Andrew E. Gelman*; Massachusetts Institute of Technology, *Richard A. Shapiro*, *Joseph L. Shipman*; Michigan State University, *Karl A. Dahlke*, *Lloyd A. Rawley*; University of New Brunswick, *Christian Friesen*; State University of New York, Buffalo, *Pierre H. Abbat*; University of Pennsylvania, *Joel Mick*; Princeton University, *Mark P. Kleiman*, *Mark A. Prysant*, *Daniel S. Rokhsar*, *Stephen A. Vavisis*; Rensselaer Polytechnic Institute, *William John Harte*; Rice University, *Christopher T. Delevoryas*, *Joe D. Warren*; Stanford University, *Thomas C. Hales*; University of Toronto, *John J. Chew*, *John J. Im*; Vanderbilt University, *Hartwig P. Arenstorf*; Washington University, St. Louis, *Bard Bloom*, *Karl F. Narveson*; University of Washington, *Kin Y. Li*; University of Waterloo, *W. Ross Brown*, *Bev I. Cope*, *Mark N. Culp*, *Herbert J. Fichtner*, *A. Michael O'Brien*; University of Wisconsin, Madison, *Chris S. Jantzen*; Yale University, *Jeffrey W. Clark*, *Alan S. Edelman*, *Paul N. Feldman*.

There were 2043 individual contestants from 343 colleges and universities in Canada and the United States in the competition of December 5, 1981. Teams were entered by 251 institutions.

The Questions Committee for the forty-second competition consisted of K. B. Stolarsky (Chairman), W. J. Firey, and D. A. Hensley; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1

Let $E(n)$ denote the largest integer k such that 5^k is an integral divisor of the product $1^1 2^2 3^3 \cdots n^n$. Calculate

$$\lim_{n\to\infty} \frac{E(n)}{n^2}.$$

Problem A-2

Two distinct squares of the 8 by 8 chessboard C are said to be adjacent if they have a vertex or side in common. Also, g is called a C-gap if for every numbering of the squares of C with all the integers $1, 2, \ldots, 64$ there exist two adjacent squares whose numbers differ by at least g. Determine the largest C-gap g.

Problem A-3

Find

$$\lim_{t\to\infty}\left[e^{-t}\int_0^t\int_0^t\frac{e^x-e^y}{x-y}\,dx\,dy\right]$$

or show that the limit does not exist.

Problem A-4

A point P moves inside a unit square in a straight line at unit speed. When it meets a corner it escapes. When it meets an edge its line of motion is reflected so that the angle of incidence equals the angle of reflection.

Let $N(T)$ be the number of starting directions from a fixed interior point P_0 for which P escapes within T units of time. Find the least constant a for which constants b and c exist such that

$$N(T) \le aT^2 + bT + c$$

for all $T > 0$ and all initial points P_0.

Problem A-5

Let $P(x)$ be a polynomial with real coefficients and form the polynomial

$$Q(x) = (x^2+1)P(x)P'(x) + x\left([P(x)]^2 + [P'(x)]^2\right).$$

Given that the equation $P(x) = 0$ has n distinct real roots exceeding 1, prove or disprove that the equation $Q(x) = 0$ has at least $2n - 1$ distinct real roots.

Problem A-6

Suppose that each of the vertices of ΔABC is a lattice point in the (x, y)-plane and that there is exactly one lattice point P in the *interior* of the triangle. The line AP is extended to meet BC at E. Determine the largest possible value for the ratio of lengths of segments

$$\frac{|AP|}{|PE|}.$$

[A lattice point is a point whose coordinates x and y are integers.]

Problem B-1

Find

$$\lim_{n\to\infty}\left[\frac{1}{n^5}\sum_{h=1}^{n}\sum_{k=1}^{n}\left(5h^4 - 18h^2k^2 + 5k^4\right)\right].$$

Problem B-2,

Determine the minimum value of

$$(r-1)^2 + \left(\frac{s}{r}-1\right)^2 + \left(\frac{t}{s}-1\right)^2 + \left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \le r \le s \le t \le 4$.

Problem B-3

Prove that there are infinitely many positive integers n with the property that if p is a prime divisor of $n^2 + 3$, then p is also a divisor of $k^2 + 3$ for some integer k with $k^2 < n$.

Problem B-4

Let V be a set of 5 by 7 matrices, with real entries and with the property that $rA + sB \in V$ whenever $A, B \in V$ and r and s are scalars (i.e., real numbers). *Prove or disprove* the following assertion: If V contains matrices of ranks 0, 1, 2, 4, and 5, then it also contains a matrix of rank 3.

[The rank of a nonzero matrix M is the largest k such that the entries of some k rows and some k columns form a k by k matrix with a nonzero determinant.]

Problem B-5

Let $B(n)$ be the number of ones in the base two expression for the positive integer n. For example, $B(6) = B(110_2) = 2$ and $B(15) = B(1111_2) = 4$. Determine whether or not

$$\exp\left(\sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)}\right)$$

is a rational number. Here $\exp(x)$ denotes e^x.

Problem B-6

Let C be a fixed unit circle in the Cartesian plane. For any convex polygon P each of whose sides is tangent to C, let $N(P, h, k)$ be the number of points common to P and the unit circle with center at (h, k). Let $H(P)$ be the region of all points (x, y) for which $N(P, x, y) \geq 1$ and $F(P)$ be the area of $H(P)$. Find the smallest number u with

$$\frac{1}{F(P)} \iint N(P, x, y)\, dx\, dy < u$$

for all polygons P, where the double integral is taken over $H(P)$.

In the 12-tuple $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, n_i for $10 \geq i \geq 0$ is the number of students among the top 209 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

SOLUTIONS

A-1. (33, 16, 13, 6, 0, 0, 0, 5, 23, 65, 39, 9)

We show that the limit is 1/8. Let $T(m) = 1 + 2 + \cdots + m = m(m+1)/2$, $[x]$ denote the greatest integer in x, $h = [\log_5 n]$, and e_i be the fractional part $(n/5^i) - [n/5^i]$ for $1 \leq i \leq h$. Then

$$E(n) = 5T([n/5]) + 5^2T([n/5^2]) + \cdots + 5^hT([n/5^h])$$

$$2E(n) = 5([n/5]^2 + [n/5]) + 5^2([n/5^2]^2 + [n/5^2]) + \cdots + 5^h([n/5^h]^2 + [n/5^h])$$

$$= 5\left(\frac{n^2}{5^2} - \frac{2e_1n}{5} + e_1^2 + \frac{n}{5} - e_1\right) + \cdots + 5^h\left(\frac{n^2}{5^{2h}} - \frac{2e_hn}{5^h} + e_h^2 + \frac{n}{5^h} - e_h\right)$$

$$\frac{E(n)}{n^2} = \frac{1}{2}\left(\frac{1}{5} + \frac{1}{5^2} + \cdots + \frac{1}{5^h}\right) + \frac{h}{2n} - \frac{e_1 + e_2 + \cdots + e_h}{n}$$

$$+ \frac{5(e_1^2 - e_1) + \cdots + 5^h(e_h^2 - e_h)}{2n^2}.$$

Since $5^h \leq n < 5^{h+1}$ and $0 \leq e_i < 1$, one sees that $h/n \to 0$ and $E(n)/n^2 \to 1/8$ as $n \to \infty$.

A-2. (86, 17, 4, 0, 1, 1, 0, 0, 28, 2, 24, 46)

For any numbering, one can go from the square numbered 1 to the square numbered 64 in 7 or fewer steps, in each step going to an adjacent square; thus $(64 - 1)/7 = 9$ is a C-gap. It is the largest C-gap since with coordinates (a, b), $1 \leq a \leq 8$ and $1 \leq b \leq 8$, for the squares we can number (a, b) with $8(a - 1) + b$ and thus find that no number greater than 9 is a C-gap.

A-3. (1, 0, 5, 0, 0, 0, 0, 0, 4, 14, 64, 121)

Let $G(t)$ be the double integral. Then

$$\lim_{t\to\infty}\left[G(t)/e^t\right] = \lim_{t\to\infty}\left[G'(t)/e^t\right]$$

by L'Hôpital's Rule. One finds that

$$G'(t) = \int_0^t \frac{e^x - e^t}{x - t}\,dx + \int_0^t \frac{e^y - e^t}{y - t}\,dy = 2\int_0^t \frac{e^x - e^t}{x - t}\,dx.$$

Then using $e^x = e^t[1 + (x - t) + (x - t)^2/2! + \cdots]$, one sees that $e^{-t}G'(t) \to \infty$ as $t \to \infty$ since for sufficiently large t,

$$\frac{G'(t)}{2e^t} = \int_0^t \frac{e^{x-t} - 1}{x - t}\,dx = \int_0^t \frac{1 - e^{-y}}{y}\,dy > \int_1^t \frac{1 - e^{-y}}{y}\,dy > (1 - e^{-1})\log t.$$

A-4. (1, 10, 9, 0, 0, 0, 0, 20, 37, 17, 21, 94)

Set up coordinates so that a vertex of the given unit square is (0, 0) and two sides of the square are on the axes. Using the reflection properties, one can see that P escapes within T units of time if and only if the (infinite) ray from P_0, with the direction of the first segment of the path, goes through a lattice point (point with integer coordinates) within T units of distance from P_0. Thus $N(T)$ is at most the number $L(T)$ of lattice points in the circle with center at P_0 and radius T. Tiling the plane with unit squares having centers at the lattice points and considering areas, one sees that

$$N(T) \le L(T) \le \pi\left[T + (\sqrt{2}/2)\right]^2.$$

Hence there is an upper bound for $N(T)$ of the form $\pi T^2 + bT + c$, with b and c fixed. When just one coordinate of P_0 is irrational,

$$N(T) = L(T) \ge \pi\left[T - (\sqrt{2}/2)\right]^2.$$

This lower bound for $N(T)$ exceeds $aT^2 + bT + c$ for sufficiently large T if $a < \pi$; hence π is the desired a.

A-5. (3, 3, 9, 0, 0, 0, 0, 0, 10, 6, 49, 129)

We show that $Q(x)$ has at least $2n - 1$ real zeros. One finds that $Q(x) = F(x)G(x)$, where

$$F(x) = P'(x) + xP(x) = e^{-x^2/2}\left[e^{x^2/2}P(x)\right]',\; G(x) = xP'(x) + P(x) = [xP(x)]'.$$

We can assume that $P(x)$ has exactly n zeros a_i exceeding 1 with $1 < a_1 < a_2 < \cdots < a_n$. It follows from Rolle's Theorem that $F(x)$ has $n - 1$ zeros b_i and $G(x)$ has n zeros c_i with

$$1 < a_1 < b_1 < a_2 < b_2 < \cdots < b_{n-1} < a_n, \quad 0 < c_1 < a_1 < c_2 < a_2 < \cdots < c_n < a_n.$$

If $b_i \neq c_{i+1}$ for all i, the b's and c's are $2n - 1$ distinct zeros of $Q(x)$. So we assume that $b_i = c_{i+1} = r$ for some i. Then

$$P'(r) + rP(r) = 0 = rP'(r) + P(r)$$

and so $(r^2 - 1)P(r) = 0$. Since $r = b_i > 1$, $P(r) = 0$. Since $a_i < r < a_{i+1}$, this contradicts the fact that the a's are all the zeros exceeding 1 of $P(x)$. Hence $Q(x)$ has at least $2n - 1$ distinct real zeros.

A-6. (1, 0, 1, 1, 0, 0, 0, 0, 17, 9, 45, 135)

Treating each point X of the plane as the vector $\overrightarrow{AX}$ with initial point at A and final point at X, let

$$L = (B + C)/2, M = C/2, \quad \text{and} \quad N = B/2$$

(be the midpoints of sides BC, AC, and AB). Also let

$$S = (2L + M)/3 = (B + C + M)/3, T = (2L + N)/3$$
$$= (B + C + N)/3, Q = 2P - B, \quad \text{and} \quad R = 3P - B - C.$$

Clearly Q and R are lattice points. Also $Q \neq P$ and $R \neq P$ since $Q = P$ implies $P = B$ and $R = P$ implies that P is the point L on side BC. Hence Q is not inside ΔABC and this implies that P is not inside ΔNBL since the linear transformation f with $f(X) = 2X - B$ translates a doubled ΔNBL (and its inside) onto ΔABC (and its inside). Similarly, P is not inside ΔMCL. Using the mapping $g(X) = 3X - B - C$ and the fact that R is not inside ΔLMN, one finds that P is not inside ΔLST. Since the distance from A to line ST is 5 times the distance between lines ST and BC, it follows that $|AP|/|PE| \leq 5$. This upper bound 5 is seen to be the maximum by considering the example with $A = (0,0)$, $B = (0,2)$, and $C = (3,0)$ in which $P = (1,1) = T$ is the only lattice point inside ΔABC and $|AT|/|TE| = 5$.

B-1. (78, 19, 33, 0, 5, 0, 0, 0, 15, 11, 33, 15)

Let $S_k(n) = 1^k + 2^k + \cdots + n^k$. Using standard methods of calculus texts one finds that

$$S_2(n) = (n^3/3) + (n^2/2) + an$$

and

$$S_4(n) = (n^5/5) + (n^4/2) + bn^3 + cn^2 + dn,$$

with a, b, c, d constants. Then the double sum is

$$10nS_4(n) - 18[S_2(n)]^2 = (2n^6 + 5n^5 + \cdots) - (2n^6 + 6n^5 + \cdots) = -n^5 + \cdots$$

and the desired limit is -1.

B-2. (37, 1, 10, 0, 0, 0, 0, 0, 6, 9, 123, 23)

First we let $0 < a < b$ and seek the x that minimizes

$$f(x) = \left(\frac{x}{a} - 1\right)^2 + \left(\frac{b}{x} - 1\right)^2 \text{ on } a \leq x \leq b.$$

Let $x/a = z$ and $b/a = c$. Then

$$f(x) = g(z) = (z - 1)^2 + \left(\frac{c}{z} - 1\right)^2.$$

Now $g'(z) = 0$ implies

$$z^4 - z^3 + cz - c^2 = (z^2 - c)(z^2 - z + c) = 0;$$

the only positive solution is $z = \sqrt{c}$. Since $0 < a < b$, $c > 1, \sqrt{c} > 1$, and

$$g(1) = g(c) = (c - 1)^2 = (\sqrt{c} - 1)^2(\sqrt{c} + 1)^2 > 2(\sqrt{c} - 1)^2 = g(\sqrt{c}).$$

Hence the minimum of $g(z)$ on $1 \leq z \leq c$ occurs at $z = \sqrt{c}$. It follows that the minimum for $f(x)$ on $a \leq x \leq b$ occurs at $x = a\sqrt{b/a} = \sqrt{ab}$. Then the minimum for the given function of r, s, t occurs with $r = \sqrt{s}$, $t = \sqrt{4s} = 2r$, and $s = \sqrt{rt} = r\sqrt{2}$. These imply that $r = \sqrt{2}$, $s = 2$, $t = 2\sqrt{2}$. Thus the desired minimum value is $4(\sqrt{2} - 1)^2 = 12 - 8\sqrt{2}$.

B-3. (5, 4, 1, 0, 0, 0, 0, 0, 3, 0, 56, 140)

As m ranges through all nonnegative integers,

$$n = (m^2 + m + 2)(m^2 + m + 3) + 3$$

takes on an infinite set of positive integral values. Let $f(x) = x^2 + 3$. Examination of $\{f(m)\} = 3, 4, 7, 12, 17, 28, 39, 52, 67, 84, \ldots$ leads one to conjecture that

$$f(x)f(x+1) = f[x(x+1)+3] = f(x^2+x+3).$$

This is easily proved. Using this property and the above relation between m and n, one has

$$f(n) = f(m^2+m+2)f(m^2+m+3) = f(m^2+m+2)f(m)f(m+1).$$

Thus $p|f(n)$ with p prime implies that $p|f(k)$ with k equal to m, $m+1$, or m^2+m+2. Since each of these possibilities for k satisfies $k^2 < n$, the desired result follows.

B-4. (38, 4, 3, 1, 0, 0, 0, 1, 3, 6, 28, 125)

Let $M = M(a, b, c)$ denote the 5 by 7 matrix (a_{ij}) with

$$a_{11} = a,\ a_{22} = a_{33} = a_{44} = a_{55} = b,\ a_{16} = a_{27} = c,$$

and $a_{ij} = 0$ in all other cases. Then the set V of all such M (with a, b, c arbitrary real numbers) is closed under linear combinations. Also, $M(0,0,0)$, $M(1,0,0)$, $M(0,0,1)$, $M(0,1,0)$, and $M(1,1,0)$ have ranks 0, 1, 2, 4, and 5, respectively. But no M in V has rank 3 since $b \neq 0$ implies that the rank is 4 or 5 and $b = 0$ forces the rank to be 0, 1, or 2.

B-5. (7, 10, 7, 6, 3, 1, 1, 3, 2, 6, 49, 114)

If n has d digits in base 2, $2^{d-1} \le n$ and so

$$B(n) \le d \le 1 + \log_2 n.$$

This readily implies that $\sum_{n=1}^{\infty}[B(n)/n(n+1)]$ converges to a real number S. Hence the manipulations below with convergent series are allowable in the two solutions which follow.

Each n is uniquely expressible as $n_0 + 2n_1 + 2^2 n_2 + \cdots$ with each n_i in $\{0, 1\}$ (and with $n_i = 0$ for all but a finite set of i). Since

$$1 + 2 + 2^2 + \cdots + 2^{i-1} = 2^i - 1,$$

one sees that $n_i = 1$ if and only if n is of the form $k + 2^i + 2^{i+1}j$ with k in $\{0, 1, \ldots, 2^i - 1\}$ and j in $\{0, 1, 2, \ldots\}$. Thus

$$\begin{aligned} S &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \sum_{i=0}^{\infty} n_i \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{2^i-1} \frac{1}{(k + 2^i + 2^{i+1}j)(1 + k + 2^i + 2^{i+1}j)}. \end{aligned}$$

Using $1/s(s+1) = 1/s - 1/(s+1)$, the innermost sum telescopes and

$$S = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[\frac{1}{2^i(1+2j)} - \frac{1}{2^i(2+2j)}\right] = \sum_{i=0}^{\infty} \frac{1}{2^i} \sum_{j=0}^{\infty} (-1)^j \frac{1}{j}.$$

Since it is well known that $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln 2$,

$$S = \left(\sum_{i=0}^{\infty} 2^{-i}\right) \ln 2 = 2 \ln 2 = \ln 4$$

and e^S is the rational number 4.

Alternatively, we note that $B(2m) = B(m)$, $B(2m+1) = 1 + B(2m) = 1 + B(m)$.

Then

$$S = \sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)} = \sum_{m=0}^{\infty} \frac{B(2m+1)}{(2m+1)(2m+2)} + \sum_{m=1}^{\infty} \frac{B(2m)}{2m(2m+1)}$$

$$= \sum_{m=0}^{\infty} \frac{1+B(m)}{(2m+1)(2m+2)} + \sum_{m=1}^{\infty} \frac{B(m)}{2m(2m+1)}$$

$$= \sum_{m=0}^{\infty} \frac{1}{(2m+1)(2m+2)} + \sum_{m=1}^{\infty} B(m)\left[\frac{1}{2m(2m+1)} + \frac{1}{(2m+1)(2m+2)}\right]$$

$$= \ln 2 + \frac{1}{2}\sum_{m=1}^{\infty} \frac{B(m)}{m(m+1)} = \ln 2 + \frac{S}{2}.$$

Hence $S/2 = \ln 2$, $S = \ln 4$, and $\exp(S)$ is the rational number 4.

B-6. (10, 0, 2, 0, 0, 0, 0, 0, 0, 7, 52, 138)

Let $L = L(P)$ be the perimeter of P. One sees that $H(P)$ consists of the region bounded by P, the regions bounded by rectangles whose bases are the sides of P and whose altitudes equal 1, and sectors of unit circles which can be put together to form one unit circle. Hence

$$F(P) = (L/2) + L + \pi = \pi + 3L/2.$$

If A and B are two consecutive vertices of P, the contribution of side AB to the double integral I is double the area of the region (of the figure) bounded by the unit semicircles with centers at A and B and segments CD and EF such that $ABCD$ and $ABEF$ are rectangles and $|AD| = 1 = |AF|$.

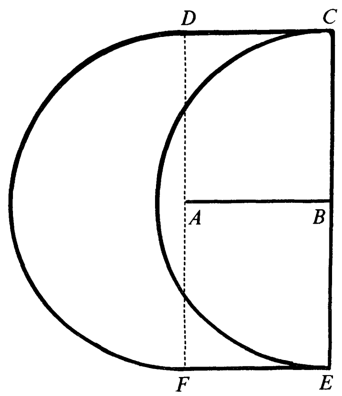


FIG. 1.

One doubles this area because there is a symmetric region bounded by CD, EF, and the other halves of the unit circles centered at A and B. The overlap of the two regions counts twice. By Cavalieri's slicing principle, this contribution of side AB to I is 4 times the length of AB. Hence $I = 4L$ and

$$\frac{I}{F(P)} = \frac{4L}{\pi + 3L/2} = \frac{8}{3 + (2\pi/L)}.$$

One can make L arbitrarily large (e.g., by letting P be a triangle with two angles arbitrarily close to right angles). Hence the desired least upper bound is $8/3$.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

A CONVEXITY PROOF OF HADAMARD'S INEQUALITY

A. W. MARSHALL
Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada V6T1X4

I. OLKIN
Department of Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305

However unlikely the discovery of yet another proof of Hadamard's inequality may be, the proof here seems to be new. It is particularly interesting to find that Hadamard's inequality follows from Jensen's inequality via a convexity result of Minkowski.

Hadamard's Inequality states that if $H = (h_{ij})$ is an $n \times n$ positive semidefinite Hermitian matrix, then

$$\det H \leqslant \prod_1^n h_{ii}. \tag{1}$$

Indeed, since any positive semidefinite Hermitian matrix H has a representation $H = CC^*$, where $C = (c_{ij})$ is an $n \times n$ complex matrix, (1) can be rewritten in the more familiar form

$$|\det C|^2 \leqslant \prod_{i=1}^n \sum_{\alpha=1}^n |c_{i\alpha}|^2.$$

Jensen's Inequality states that if ϕ is an extended real valued convex function defined on a convex set A of $n \times n$ matrices, $A_1, \ldots, A_k \in A$ and $\alpha_1, \ldots, \alpha_k$ are nonnegative numbers satisfying $\Sigma\alpha_i = 1$, then

$$\phi(\alpha_1 A_1 + \cdots + \alpha_k A_k) \leqslant \alpha_1\phi(A_1) + \cdots + \alpha_k\phi(A_k). \tag{2}$$

Finally, a *Convexity Result of Minkowski* states that if H is an $n \times n$ positive semidefinite Hermitian matrix, then (with the convention $\log 0 = -\infty$)

$$\log\det H \text{ is concave,} \tag{3}$$

$$(\det H)^{1/n} \text{ is concave.} \tag{4}$$

(See e.g., Bellman (1970), p. 128, 132, Marcus and Minc (1964), p. 115, or Marshall and Olkin (1979), p. 475, 476.) Note that (4) implies (3) by virtue of the fact that concavity implies logconcavity.

To prove Hadamard's Inequality, let

$$H_i = D_i H D_i, \quad i = 1, \ldots, 2^n,$$

where the D_i are distinct diagonal matrices of the form $D_i = \operatorname{diag}(\varepsilon_1, \ldots, \varepsilon_n)$, with $\varepsilon_j = \pm 1$, $j = 1, \ldots, n$. Then

$$\det H_i = (\det D_i)^2 (\det H) = \det H. \tag{5}$$

The key to this proof is the observation that for $\alpha_i = 1/2^n$, $i = 1, \ldots, 2^n$,

$$\Sigma\alpha_i H_i = \Sigma\alpha_i D_i H D_i = \operatorname{diag}(h_{11}, \ldots, h_{nn}). \tag{6}$$

With the choice $\phi(A) = -\log\det A$, (2) yields

$$\log\prod_1^n h_{ii} = \log\det(\Sigma\alpha_i H_i) \geqslant \Sigma\alpha_i \log\det H_i = \log\det H. \tag{7}$$

With the choice $\phi(A) = -(\det A)^{1/n}$, (2) yields

$$(\Pi h_{ii})^{1/n} = (\det\Sigma\alpha_i H_i)^{1/n} \geqslant \Sigma\alpha_i(\det H_i)^{1/n} = (\det H)^{1/n}. \tag{8}$$

Both (7) and (8) are Hadamard's Inequality, although (8) makes use of a stronger concavity property.

Also obtainable using the same approach is the following *Result of Schur (1923)*: If H is an $n \times n$ Hermitian matrix with diagonal elements $h_{11},\ldots,h_{nn}$ ordered $h_{11} \geqslant \cdots \geqslant h_{nn}$ and with characteristic roots $\lambda_1(H),\ldots,\lambda_n(H)$ ordered $\lambda_1(H) \geqslant \cdots \geqslant \lambda_n(H)$, then

$$h_{11} + \cdots + h_{kk} \leqslant \lambda_1(H) + \cdots + \lambda_k(H), \quad k = 1,\ldots,n \tag{9}$$

with equality for $k = n$.

To prove (9) use the fact that $\phi(A) = \lambda_1(A) + \cdots + \lambda_k(A)$ is convex on the set of Hermitian matrices for $k = 1,\ldots,n$. (See e.g., Bellman (1970), p. 132, or Marshall and Olkin (1979), p. 478.) Consequently, with the same choice of H_i and α_i, $i = 1,\ldots,2^n$, as in the previous proof,

$$\sum_{j=1}^k h_{jj} = \sum_{j=1}^k \lambda_j(\operatorname{diag}(h_{11},\ldots,h_{nn})) = \sum_{j=1}^k \lambda_j(\Sigma\alpha_i H_i)$$

$$\leqslant \Sigma\alpha_i \sum_{j=1}^k \lambda_j(H_i) = \Sigma\alpha_i \sum_{j=1}^k \lambda_j(H) = \sum_1^k \lambda_j(H),$$

$k = 1,\ldots,n$. Equality for $k = n$ is immediate from $\operatorname{tr} H = \Sigma_1^n \lambda_i(H) = \Sigma_1^n h_{ii}$.

The inequalities (9) define a partial ordering of vectors. It was already noted by Schur (1923) that this partial ordering implies Hadamard's Inequality.

Acknowledgement. The authors are grateful to Professor T. Cover for the conversations concerning a proof of Hadamard's inequality from the information inequality. This work was supported in part by the National Science Foundation.

References

1. R. Bellman, Introduction to Matrix Analysis, 2nd ed., McGraw-Hill, New York, 1970.
2. M. Marcus and H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Boston, 1964.
3. A. W. Marshall and I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, New York, 1979.
4. I. Schur, Über eine Klasse von Mittelbildungen mit Anwendungen die Determinantan, Theorie Sitzungsber. Berlin. Math. Gesellschaft, 22 (1923) 9–20.

GENERALIZED HYPERBOLIC FUNCTIONS

ABRAHAM UNGAR*
Department of Mathematics, University of California, Berkeley, CA 94720

Generalized hyperbolic functions, appearing to have a compelling intrinsic beauty, are presented.

For any pair of integers (n, r), $n \geqslant 2$, $0 \leq r \leq n - 1$, let the entire function in the complex z-plane C,

*Permanent address: Dept. of Mathematics, Rhodes University, Grahamstown, South Africa.

$$H_{n,r}(z)=\sum_{k=0}^{\infty}\frac{z^{nk+r}}{(nk+r)!} \tag{1}$$

be referred to as the hyperbolic function of order n and rth kind. The two hyperbolic functions of order $n=2$ are thus $H_{2,0}(z)=\cosh z$ and $H_{2,1}(z)=\sinh z$.

Since

$$\frac{d}{dz}H_{n,r}(z)=H_{n,r-1}(z), \tag{2}$$

where we define

$$H_{n,-1}(z)=H_{n,n-1}(z), \tag{3}$$

the hyperbolic functions of order n,

$$H_{n,r}(z),\quad r=0,1,2,\ldots,(n-1),$$

form a set of n linearly independent solutions of the differential equation

$$\frac{d^n\phi(z)}{dz^n}=\phi(z). \tag{4}$$

Furthermore, let us extend the definition of $H_{n,r}(z)$ to any integer r,

$$H_{n,r}(z)=H_{n,r(\mathrm{mod}\ n)}(z), \tag{5}$$

and let $H_n(z)$ be the $n\times n$ matrix

$$H_n(z)=(a_{ij})$$

where

$$a_{ij}=H_{n,i-j}(z),\quad i,j=0,1,2,\ldots,(n-1).$$

The matrix $H_n(z)$, $n\geqslant 2$, referred to as the hyperbolic matrix of order n, is a matrix of very special type, a so-called *circulant matrix*,

$$H_n(z)=\begin{pmatrix} H_{n,0}(z) & H_{n,n-1}(z) & H_{n,n-2}(z) & \cdots & H_{n,1}(z)\\ H_{n,1}(z) & H_{n,0}(z) & H_{n,n-1}(z) & \cdots & H_{n,2}(z)\\ H_{n,2}(z) & H_{n,1}(z) & H_{n,0}(z) & \cdots & H_{n,3}(z)\\ \vdots & & & & \\ H_{n,n-1}(z) & H_{n,n-2}(z) & H_{n,n-3}(z) & \cdots & H_{n,0}(z)\end{pmatrix},\quad z\in C. \tag{6}$$

A circulant matrix is an $n\times n$ matrix in which the first column consists of n distinct elements and the successive columns are the successive cyclic permutations of these. Circulant matrices have recently become of increased interest; see [**1**], [**2**].

For $n=1$, $H_{1,0}(z)=e^z$ is the exponential function and so is the 1×1 matrix $H_1(z)$ whose only element is $H_{1,0}(z)$.

THEOREM.

$$|H_n(z)|=1,\quad z\in C, \tag{7}$$

and

$$H_n(z_1+z_2)=H_n(z_1)\cdot H_n(z_2),\quad z_1,z_2\in C, \tag{8}$$

where $|H_n(z)|$ *is the determinant of the matrix* $H_n(z)$, $n\geqslant 2$, *and* $\cdot$ *assigns matrix multiplication.*

For $n=2$, the two identities in the Theorem specialize to

$$\begin{vmatrix} \cosh z & \sinh z \\ \sinh z & \cosh z \end{vmatrix} = 1 \tag{7'}$$

and

$$\begin{pmatrix} \cosh(z_1 + z_2) & \sinh(z_1 + z_2) \\ \sinh(z_1 + z_2) & \cosh(z_1 + z_2) \end{pmatrix} = \begin{pmatrix} \cosh z_1 & \sinh z_1 \\ \sinh z_1 & \cosh z_1 \end{pmatrix} \begin{pmatrix} \cosh z_2 & \sinh z_2 \\ \sinh z_2 & \cosh z_2 \end{pmatrix}. \tag{8'}$$

The Theorem indicates that the circulant matrix $H_n(z)$, $n \geqslant 2$, $z \in C$, represents a group with the group multiplication given by addition of the argument z. For $n = 2$ and $z = \alpha$ real, the group represented by the hyperbolic matrix $H_2(\alpha)$ is the well-known group of pseudo-rotations, or Lorentz transformations, of the Lorentz relativistic plane. Like the exponential function $H_1(z) = e^z$, the hyperbolic matrices $H_n(z)$, $n \geqslant 2$, are never singular and take addition to (matrix) multiplication.

Proof of the Theorem. The proof of both identities (7) and (8) is obtained by employing the differentiation rule (2). Being a sum of determinants each of which has two equal rows

$$\frac{d}{dz}|H_n(z)| = 0; \tag{9}$$

and it can readily be shown that

$$\frac{d}{dz}[H_n(a + z) \cdot H_n(-z)] = 0 \tag{10}$$

for $n \geqslant 2$, $z \in C$ and any complex constant a, where the derivative of a matrix in (10) is a matrix of derivatives. To see this easily, the reader may write out the identities (9) and (10) in full for $n = 3$.

Eq. (9) implies that $|H_n(z)|$ is a constant, and the constant is unity since $H_n(0)$ is the unit $n \times n$ matrix I_n.

Eq. (10) implies that

$$H_n(a + z) \cdot H_n(-z) = H_n(a), \tag{11}$$

which upon letting $a = z_1 + z_2$ and $z = -z_2$ yields (8).

The hyperbolic functions of order n, $n \geqslant 2$, enjoy two basic properties. The first property is that they form a set of n linearly independent solutions of an ordinary differential equation that are obtained from one another by differentiations. The second property is that they form a continuous commutative group represented by an $n \times n$ matrix of a single complex variable the determinant of which is unity.

As an application to the theory of ordinary differential equations, let us consider the particular case of $n = 3$. The three hyperbolic functions of order 3, $H_{3,0}(x)$, $H_{3,1}(x)$ and $H_{3,2}(x)$, $-\infty < x < \infty$, may be written in terms of the exponential function,

$$\begin{aligned} H_{3,0}(x) &= \tfrac{1}{3}(e^x + e^{q_1 x} + e^{q_2 x}) \\ H_{3,1}(x) &= \tfrac{1}{3}(e^x + q_1 e^{q_1 x} + q_2 e^{q_2 x}) \\ H_{3,2}(x) &= \tfrac{1}{3}(e^x + q_2 e^{q_1 x} + q_1 e^{q_2 x}), \end{aligned} \tag{12}$$

where $q_1 = (-1 + i\sqrt{3})/2$ and $q_2 = (-1 - i\sqrt{3})/2$ are the two nonreal cube roots of unity.

Equivalently, the hyperbolic functions of order 3 can be written as

$$\begin{aligned} H_{3,0}(x) &= \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\tfrac{\sqrt{3}}{2}x)] \\ \mathrm{H}_{3,1}(x) &= \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\tfrac{\sqrt{3}}{2}x - 2\tfrac{\pi}{3})] \end{aligned} \tag{13}$$

$$H_{3,2}(x) = \tfrac{1}{3}[e^x + 2e^{-x/2}\cos(\frac{\sqrt{3}}{2}x + 2\frac{\pi}{3})],$$

thus indicating that the ordinary differential equation

$$\frac{d^3\phi(x)}{dx^3} = \phi(x) \tag{14}$$

has three linearly independent solutions, $H_{3,r}(x)$, $r = 0,1,2$, approaching $e^x/3$ as $x \to \infty$, and rapidly oscillating near $x = -\infty$.

Acknowledgement. Calculations in the present work were performed by employing the symbolic manipulation system VAXIMA while visiting Professor Richard J. Fateman at the University of California, Berkeley, to whom the author is indebted. Work reported herein was supported in part by the U.S. Department of Energy, Contract DE-AT03767SF00034, Project Agreement DE-AS03-79ER10358, awarded to R. J. Fateman. The author wishes to thank a referee for a useful suggestion.

References

1. P. J. Davis, Circulant Matrices, Wiley, New York, 1979.
2. E. T. Wong, Polygons, circulant matrices, and Moore-Penrose inverses, this MONTHLY, 88(1981) 509–515.

VAN DER WAERDEN'S CONTINUOUS NOWHERE DIFFERENTIABLE FUNCTION

PATRICK BILLINGSLEY
Department of Mathematics, The University of Chicago, Chicago IL 60637

If f has derivative (two-sided) $f'(x)$ at x, if $u_n \leqslant x \leqslant v_n$ and $u_n < v_n$, and if $v_n - u_n \to 0$, then

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} \to f'(x).$$

Van der Waerden's function is $f(x) = \sum_{k=0}^{\infty} a_k(x)$, where $a_0(x)$ is the distance from x to the nearest integer and $a_k(x) = 2^{-k}a_0(2^k x)$. By the Weierstrass M-test, f is continuous. If u is a dyadic rational of order n (has the form $i2^{-n}$), then $2^k u$ is an integer for $k \geqslant n$, and so $f(u) = \sum_{k=0}^{n-1} a_k(u)$. Let u_n and v_n be the pair of successive dyadic rationals of order n ($v_n - u_n = 2^{-n}$) for which $u_n \leqslant x < v_n$ (x fixed). Then

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} = \sum_{k=0}^{n-1} \frac{a_k(v_n) - a_k(u_n)}{v_n - u_n}.$$

Since a_k is linear over $[u_n, v_n]$ for $0 \leqslant k < n$, the difference quotient in the sum is the right-hand derivative $a_k^+(x)$. But since $a_k^+(x) = \pm 1$,

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} = \sum_{k=0}^{n-1} a_k^+(x)$$

cannot converge to a finite limit.

This argument makes precise the idea that if f were differentiable, the impossible equation

$$f'(x) = \sum_{k=0}^{\infty} a_k'(x) = \sum_{k=0}^{\infty} \pm 1$$

ought to follow. The usual argument in effect involves the base 10 instead of the dyadic expansion; it is less transparent than the proof above because it defines u_n and v_n in a complicated way to ensure that $u_n = x$ or $v_n = x$ (and $v_n - u_n = 10^{-n}$) for each n.

ON TRISECTION, QUINTISECTION, . . . , ETC.

DONALD DUNCAN
Department of Mathematics, Sonoma State University, Rohnert Park, CA 94928

WILLIAM BARNIER
Department of Mathematics, Sonoma State University, Rohnert Park, CA 94928

Many courses in Modern Algebra prove that a real number a can be constructed with ruler and compass only if the field extension $Q(a)$ of the rationals Q has degree 2^n, i.e., $[Q(a):Q] = 2^n$, see [**1**].

The standard proof that an arbitrary angle cannot be trisected with ruler and compass is done by showing that 60° cannot be trisected by noting that the cubic equation

$$4(\cos a)^3 - 3\cos a = 1/2 = \cos 60°$$

is irreducible over the rationals and hence $[Q(\cos 20°):Q] = 3 \neq 2^n$.

The following short proof shows that for an odd prime p an arbitrary angle cannot be p-sected. We do this by finding an infinite class of angles $\beta = p\alpha$ for which α cannot be constructed. (Incidentally the β are easily constructed.)

Consider the infinite class of angles: $\beta = p\alpha = \cos^{-1}(p/N)$ with $N > p$ and $(N, p) = 1$. Now, see Figure 1,

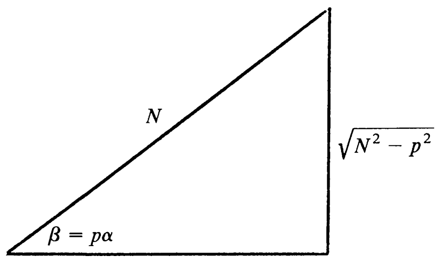


FIG. 1.

$$(\cos\alpha + i\sin\alpha)^p = \cos p\alpha + i\sin p\alpha = \frac{p}{N} + \frac{i\sqrt{N^2 - p^2}}{N}.$$

Equating real parts we have

$$(\cos\alpha)^p - \frac{p(p-1)}{2}(\cos\alpha)^{p-2}(1 - \cos^2\alpha) + \cdots = \frac{p}{N}.$$

Noting that the binomial coefficients $\binom{p}{n}$ are divisible by p (except for $\binom{p}{p}$ which is not involved), multiplying through by N and collecting coefficients of the powers of $\cos\alpha$, we obtain:

$$(N + k_1 p)(\cos\alpha)^p + k_2 p(\cos\alpha)^{p-2} + \cdots + k_{(p+1)/2}\, p(\cos\alpha) - p = 0.$$

Since $(N, p) = 1$, this polynomial is irreducible by Eisenstein's criteria. Hence we have

$$[Q(\cos\alpha):Q] = p \neq 2^n$$

and so $\cos\alpha$ and hence α cannot be constructed.

Reference

1. Charles R. Hadlock, Field Theory and Its Classical Problems, Carus Mathematical Monograph No. 19, The Mathematical Association of America, Washington, D.C., 1979.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

SET THEORY AND THE INDICATOR FUNCTION

JOHN E. MORRILL
Department of Mathematics, DePauw University, Greencastle, IN 46135

Instructors who are faced with the problem of selecting a teaching method for the verification of set-theoretic statements in such courses as finite mathematics or elementary probability theory usually find that they have several choices: (1) give rigorous arguments which use the definitions of union, intersection, equality, inclusion, etc.; (2) use Venn diagrams and do a lot of "arm-waving"; (3) show that the statements hold for a few specific small finite sets; (4) don't verify the statements in any fashion.

Consider the problem of proving that two sets, A and B, are equal. As a consequence of the definition of set equality we have to show that if $x \in A$, then $x \in B$, and if $x \in B$, then $x \in A$; or equivalently for the latter, if $x \notin A$, then $x \notin B$. In any event we are essentially concerned with examining two cases—the case where $x \in A$ and the case where $x \notin A$. In-out, yes-no, on-off,.... Of course. Such a situation suggests the introduction here of the characteristic function or what some prefer to call the indicator function.

DEFINITION: If U is the universe set and A is any subset of U, the indicator function on U associated with A is

$$I_A(x) = \begin{cases} 0, & \text{for } x \notin A \\ 1, & \text{for } x \in A. \end{cases}$$

DEFINITION: $I_A = I_B$ if and only if $I_A(x) = I_B(x)$ for all $x \in U$.

The proofs of the following theorems will be omitted since they are easy. However, we shall give some examples of their use.

THEOREM 1: (a) $I_U = 1$, *where the* 1 *here is the function on U which is identically* 1.
(b) $I_\varnothing = 0$, *where the* 0 *here is the function on U which is identically* 0.
(c) $I_A I_A = I_A$.
(d) $I_A + I_B = I_B + I_A$.
(e) $I_A I_B = I_B I_A$.
(f) $I_{A \cap B} = I_A I_B$.
(g) $I_{A \cup B} = I_A + I_B - I_A I_B$.
(h) $I_{A'} = 1 - I_A$, *where A' is the complement of A.*

THEOREM 2: $I_A = I_B$ *if and only if* $A = B$.

DEFINITION: $I_A \leqslant I_B$ if and only if $I_A(x) \leqslant I_B(x)$ for all $x \in U$, and $I_A < I_B$ if and only if $I_A \leqslant I_B$ and there is at least one $x \in U$ such that $I_A(x) = 0$ and $I_B(x) = 1$.

THEOREM 3: (a) $I_A I_B \leqslant I_A$.
(b) $I_A \leqslant I_A + I_B$.
(c) If $I_A \leqslant I_B$ *and* $I_B \leqslant I_C$, *then* $I_A \leqslant I_C$.

THEOREM 4: (a) $I_A \leqslant I_B$ *if and only if* $A \subseteq B$.
(b) $I_A < I_B$ *if and only if* $A \subset B$.

Now let us consider a few examples. The verification of one of the distributive laws, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, follows immediately since the indicator function for each side can be written, using Theorem 1, as $I_A + I_B I_C - I_A I_B I_C$. Verification of one of the DeMorgan Laws, $(A \cup B)' = A' \cap B'$, can be done as follows:

$$I_{(A \cup B)'} = 1 - I_{A \cup B} = 1 - (I_A + I_B - I_A I_B) = (1 - I_A)(1 - I_B) = I_{A'} I_{B'} = I_{A' \cap B'}.$$

As another illustration one can show (without having to recall any previously demonstrated statements) that

$$(A \cap B') \cap (C' \cap A) = A \cap (B \cup C)'$$

from the argument

$$\begin{aligned} I_{(A \cap B') \cap (C' \cap A)} &= I_{(A \cap B')} I_{(C' \cap A)} = I_A I_{B'} I_{C'} I_A = I_A{}^2 (1 - I_B)(1 - I_C) = I_A (1 - I_B - I_C + I_B I_C) \\ &= I_A (1 - I_{B \cup C}) = I_A I_{(B \cup C)'} = I_{A \cap (B \cup C)'}. \end{aligned}$$

The method is not entirely a "mindless" one as the reader can discover by looking at how indicator functions can be used to consider such questions as: What is the relationship between $(A \cup B) \cap C$ and $A \cup (B \cap C)$? Are they equal for all A, B and C in U? Are they ever equal? Under what conditions, if any, is one a subset of the other?

In conclusion it is apparent that there is nothing astounding in what has been said above, but it is surprising that in an examination of some thirty to forty finite mathematics textbooks which included material on both functions and sets none were found which even suggested this idea. It may be that the argument against the use of the indicator function here is that it is mere "symbol manipulation," and that in part may be true. It may also be argued that this function and its arithmetic are more sophisticated than the elementary set theory to which it is applied. However, if the instructor chooses to take the time to introduce this idea, it could serve as a vehicle to provide the student with a greater appreciation and understanding of the concept of function itself (e.g., not every function is "formula" written as $y = f(x)$ which is defined on some set of numbers). In addition, it might also be good preparation for the functional concepts of probability measure and random variable. In any case, the use of indicator functions does provide a reasonably accessible method of verifying set-theoretic statements and might even give some students the pleasure of proving something in an elementary mathematics course.

WHAT EVER HAPPENED TO THE HISTORY OF MATHEMATICS?

FRANK J. SWETZ
Mathematical Sciences Program, Capitol Campus, The Pennsylvania State University, Middletown, PA 17057

A recent survey [1] of the undergraduate mathematics program at 406 collegiate institutions revealed that less than 35% of them offered a course specifically on the history of mathematics. This appears to be a rather sad commentary on the state of undergraduate mathematics teaching in the United States. While many professions, e.g., law, medicine and architecture are presently moving towards requiring studies in the history of their major discipline, the teaching of mathematics is tending more and more to divorce itself from an historical perspective. Before World War II, the history of mathematics occupied an important place in the undergraduate mathematics curriculum, particularly that of teacher training institutions, but with the explosive growth of new areas of mathematical interest and applications (statistics, linear algebra, topology, operations research, computer science, etc.) it was relegated to a position of little importance. (It is interesting to note that this has not been the case in Europe, particularly the Soviet Union, where the history of mathematics still holds a position of prominence in the university curriculum and academic research.)

Many reasons can be given for teaching a course in the history of mathematics. Some of these would be:

1. to demonstrate the development of mathematics as a necessary human activity;
2. to show that mathematical ideas evolve over a period of time, are labored upon and subject to change;
3. to expose the interrelationship of mathematics with other disciplines: anthropology, sociology, economics, politics, music, arts, etc. [**2**];
4. to develop an appreciation of the structure of mathematics as viewed from an historical perspective, and
5. to expose the interrelationship of seemingly diverse areas of mathematics [**3**]; for example, a consideration of the development for the value of pi would reveal first the use of rough geometric approximation (1000 BC), then better geometric-analytical approximation by the exhaustion method of Archimedes (200 BC), analytic results of series approximation as employed by James Gregory (1671), Comte de Buffon's probabilistic approach (1771), and finally the iterative results obtained by modern computers (1950 +).

The history of mathematics can be taught informally in all mathematics courses or formally in a specific history of mathematics offering. Historical notes and justifications can easily be injected into lectures. Anecdotes concerning problem situations from which mathematical investigations and theories evolved would seem appropriate for use in most courses; for example, the expression we now know as a Fourier series emerged from Fourier's attempts to understand heat transfer.

Since the subject contains a huge body of information, the specific topics to be studied and their sequencing must be carefully planned [**4**]. It is a mistake to try to teach the mathematical accomplishments of 5000 years in one course. A course in the history of mathematics can most beneficially be given during a student's undergraduate career either in the sophomore or senior year of study. A sophomore level offering lays before the student a panoramic view of the territory to be explored; it can serve as an introduction to the content, techniques and trends of mathematical thought. Works by Struik [**5**] and Kline [**6**] lend themselves for such purposes. A senior level course also presents an overview of mathematics, but in addition serves to summarize the students' previous studies and give human meaning to much of the material they have learned. In such efforts issues can be put into perspective: "How did the rise of industrialization stimulate the development of the study of differential equations?"; "Why was the advent of modern mercantile capitalism in 15th century Italy accompanied by a period of mathematical growth and accomplishment?"; "What effects has warfare had on mathematics?", etc. Eves' book [**7**] with its excellent selection of problem studies nicely accommodates such a course. Projects and term papers can be usefully employed to supplement classroom discussions [**8**].

Ultimately, the history of mathematics and its meaningful success as a student learning experience depends on the instructors. Many of us have had no formal exposure to this discipline. However, the history of mathematics boasts a rich literature that encourages self-study [**9**]. While the references already cited provide an introduction to the subject, more specialized treatments are available [**10**]. In particular several journals either regularly feature articles on the history of mathematics or are completely devoted to it, e.g., *Isis*, *Scripta Mathematica*, *Historia Mathematica*.

What ever happened to the history of mathematics? It's there, waiting for us to teach it.

References

1. L. A. Lechleitner, Colleges and Universities Offering Courses in the History of Mathematics, student research project, Dept. of Math. Sciences, Capitol Campus, Middletown, PA, 1979.

2. M. Kline, Mathematics a Cultural Approach, Addison-Wesley, Reading, MA, 1962.

3. R. L. Wilder, Mathematics as a Cultural System, Pergamon, New York, 1981.

4. F. J. Swetz, Some thoughts on teaching the history of mathematics, PA Council of Teachers of Math. J., 6, (1981) 9-15.

5. D. J. Struik, A Concise History of Mathematics, Dover, New York, 1967.

6. M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, New York, 1972.
7. H. Eves, An Introduction to the History of Mathematics, Holt, Rinehart and Winston, New York, 1976.
8. F. J. Swetz, The use of projects in teaching the history of mathematics, Historia Math., 9 (1982) 201–205.
9. A. R. Dorling, (Editor), The Use of Mathematical Literature, Butterworths, Boston, 1977.
10. C. Read, Periodical articles dealing with the history of mathematics: a bibliography of articles in English appearing in nine periodicals, School Science and Mathematics, 68 (1970) 415–453.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by March 31, 1983. Please place the solvers's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2968. *Proposed by George Tsintsifas, Thessaloniki, Greece.*

The points A_1', A_2', A_3' lie on the sides A_2A_3, A_3A_1, A_1A_2 of an acute angle triangle $A_1A_2A_3$, respectively. Show that

$$2\sum a_i' \cos A_i \geqslant \sum a_i \cos A_i$$

where a_1, a_2, a_3 are the sides of the triangle $A_1A_2A_3$ and a_1', a_2', a_3' are the sides of the triangle $A_1'A_2'A_3'$.

E 2969. *Proposed by John Brillhart, The University of Arizona, and Constantin Sevici, University of Michigan.*

Let $f(x) = a_0x^n + \cdots + pa_n$ be a polynomial with integer coefficients such that $a_0a_n \neq 0$, $(a_0, a_1, \ldots, pa_n) = 1$, and let p be a prime such that $p > \sum_{s=0}^{n-1} |a_s| |a_n|^{n-1-s}$. Prove that $f(x)$ is irreducible in $\boldsymbol{Z}[x]$.

E 2970. *Proposed by J. H. Conway, University of Cambridge, and B. Reznick, University of Illinois.*

Find all finite sets of points in the plane ($P = \{p_1, \ldots, p_n\}$) with the following "closure" property: If the line segments $\overline{p_i p_j}$ and $\overline{p_k p_l}$ intersect in a single point p, then $p \in P$.

E 2971. *Proposed by Barry L. Zaslove, Northeastern University.*

Find the solution curves of the ODE,

$$y''' = y'(3y''^2 - y''' y').$$

E 2972. *Proposed by J. O. Shallit, University of California, Berkeley.*

Let $p(x_1, x_2, \ldots, x_n)$ be a polynomial in n variables with constant term 0, and let $\#(p)$ denote the number of distinct terms in p after terms with like exponents have been collected. Thus for example $\#((x_1 + x_2)^5) = 6$.

Find a formula for $\#(q_n)$ where

$$q_n = x_1(x_1 + x_2)(x_1 + x_2 + x_3) \cdots (x_1 + \cdots + x_n).$$

E 2973. *Proposed by M. M. Konstantinov, Bulgaria.*

Let $(L, \leqslant)$ be a lattice, where L is a finite set and $\leqslant$ is an order relation. For $x, y \in L$ set $x \vee y = \sup(x, y)$ and $x \wedge y = \inf(x, y)$. Denote by P the set of the bijections $L \to L$. The bijection f is said to be increasing (decreasing) if $x \leqslant y$ implies $f(x) \leqslant f(y)$ ($f(x) \geqslant f(y)$). The bijection f is said to be monotone if it is either increasing or decreasing. The set P can be represented as a union $P = P^+ \cup P^- \cup P^0$ of disjoint components, where P^+, P^- and P^0 are the sets of increasing, decreasing, and nonmonotone bijections.

A. Let $f \in P$. Prove that the following three assertions are equivalent:

(i) $f(x \wedge y) = f(x) \wedge f(y)$;
(ii) $f(x \vee y) = f(x) \vee f(y)$;
(iii) $f \in P^+$.

B. Let $f \in P$. Prove that the following three assertions are equivalent:

(j) $f(x \wedge y) = f(x) \vee f(y)$;
(jj) $f(x \vee y) = f(x) \wedge f(y)$;
(jjj) $f \in P^-$.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Sequence with Variable Rules of Succession

E 2877 [1981, 209]. *Proposed by J. C. Lagarias and A. M. Odlyzko, Bell Laboratories, Murray Hill, NJ.*

Let $f(k)$ be any sequence of integers with $f(1) = 1$, $f(2) = 2$, $f(3) = 3$ and which is defined recursively by using for each $k \geqslant 4$ one of the following rules, subject to the constraint that Rule 1 can never be used twice consecutively.

Rule 1. $f(k) = f(k-1)$.
Rule 2. $f(k) = f(k-2) + f(k-3)$.

Find the largest constant θ such that $\liminf_{k\to\infty}(f(k))^{1/k} \geqslant \theta$ is valid for all such sequences $f(k)$.

Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology,

Eindhoven, the Netherlands. We shall prove that $\theta = 2^{1/3}$. Clearly the sequence defined by rule 1 for $k \equiv 2 \bmod 3$ and by rule 2 otherwise satisfies $f(3k) = f(3k+1) = f(3k+2) = 3 \cdot 2^{k-1}$. From this sequence it follows immediately that $\theta \leqslant 2^{1/3}$. To show that $\theta \geqslant 2^{1/3}$ it suffices to prove that for all k

$$(*) \qquad f(k) \geqslant \begin{cases} 2^{(k-1)/3} & \text{if } f(k) \text{ is computed by rule 2,} \\ 2^{(k-2)/3} & \text{if } f(k) \text{ is computed by rule 1.} \end{cases}$$

For $k = 1, 2, 3$ one can check $(*)$ rather trivially. If $f(k+1)$ is computed by rule 1, then $f(k)$ must have been computed by rule 2. So

$$f(k+1) = f(k) \geqslant 2^{(k-1)/3} = 2^{((k+1)-2)/3}.$$

Similarly if $f(k+1)$ is computed by rule 2 and $f(k-1)$ by rule 1 (and thus $f(k-2)$ by rule 2), then

$$f(k+1) = f(k-1) + f(k-2) = 2f(k-2) \geqslant 2 \cdot 2^{(k-3)/3} = 2^{((k+1)-1)/3}.$$

Finally if $f(k+1)$ and $f(k-1)$ are computed by rule 2, then

$$\begin{aligned} f(k+1) &= f(k-1) + f(k-2) \geqslant 2^{(k-2)/3} + 2^{(k-4)/3} \\ &= 2^{k/3}(2^{-2/3} + 2^{-4/3}) \geqslant 2^{k/3} = 2^{((k+1)-1)/3}, \end{aligned}$$

since the relation $(2^{1/3} - 2^{-1/3})^2 \geqslant 0$, implies that $2^{-2/3} + 2^{-4/3} \geqslant 1$.

Also solved by R. S. Booth (Australia), L. E. Mattics, W. A. Newcomb, Problemlösegruppe (Mannheim, Germany) and the proposers.

Number of Terms in $(x + y)^n$ mod p

E 2879 [1981, 291]. *Proposed by Paul W. Haggard, East Carolina University.*

Let D be an integral domain with characteristic p (prime). The expansion of $(x+y)^n$ (in $D[x, y]$) has $N = \Pi_{i=0}^{m}(k_i + 1)$ nonzero terms, where $n = \Sigma_{i=0}^{m} k_i p^i$ is the p-expanded form of the positive integer n. Clearly, $N \leqslant n + 1$. (a) For what integers n is $N = n + 1$? (b) Can the relation $N = n (> 0)$ hold?

Solution by Lorraine L. Foster, California State University, Northridge. (Note that $(x+y)^k$ has $k+1$ terms for $k < p$ and $(\Sigma a_i)^{p^r} = \Sigma a_i^{p^r}$ so that $N = \Pi_{i=0}^{m}(k_i + 1)$ as asserted.)

(a) $N = n + 1$ iff $k_i = p - 1$ for all $i < m$ ($k_m \neq 0$). For,

$$N = k_m \prod_{i<m} (k_i + 1) + k_{m-1} \prod_{i<m-1} (k_i + 1) + \cdots + (k_0 + 1)$$

and

$$\prod_{i<t} (k_i + 1) \leqslant p^t$$

for $2 \leqslant t \leqslant m$.

(b) $N = n$ iff $n = 2(p-1)$. For, from (a), if $N = n$, some $k_i + 1 < p$ for $i < m$ so that

$$(n+1) - N \geqslant k_m p^{m-1}(p - k_i - 1) \geqslant 2$$

unless $k_m = 1$, $m = 1$, $i = 0$, $k_0 = p - 2$, $n = 2(p-1)$. Conversely, if $n = 2(p-1)$, $N = n$.

Also solved by J. Brandler, R. Gilmer & W. Nichols, D. A. Rawsthorne, University of South Alabama Problem Group, and the proposer.

Brandler referred to I. M. Vinogradov, *Elements of number theory* (p. 138) and Dacić, Math. Mag. (Jan 1981).

A Nonhomogeneous Inequality for n Real Numbers

E 2884 [1981, 349]. *Proposed by Lawrence Harris, University of Kentucky.*

Let $x_1,\dots,x_n$ be distinct real numbers. Set $S=\Sigma_1^n(1+x_k^2)^{n/2}/P(k)$, where $P(k)=\Pi_{j\neq k}|x_k-x_j|$. Prove $S\geq n$. When does $S=n$?

Solution by I. J. Schoenberg, University of Wisconsin-Madison. We assume $n\geq 2$ and abbreviate by writing throughout $g(x)=(1+x^2)^{n/2}$. $S=S(x_1,\dots,x_n)$ is evidently a symmetric function of the x_k, and without loss of generality we may assume $x_n<x_{n-1}<\cdots<x_1$ and express S without absolute value signs as

$$S=\sum_1^n(-1)^{k-1}g(x_k)/\prod_{j\neq k}(x_k-x_j). \tag{1}$$

We also need the two curves $C_+ : y=g(x)$ and $C_- : y=-g(x)$. Using divided differences it is clear that for the polynomial $Q(x)\in\pi_{n-1}$ such that $Q(x_k)=(-1)^{k-1}g(x_k)$ for all k, we have

$$Q(x)=S(x_1,\dots,x_n)x^{n-1}+\text{lower degree terms.} \tag{2}$$

Intuitively it would seem reasonable that $S(x_1,\dots,x_n)$ should be least when the graph of $y=Q(x)$ is tangent to C_+ and to C_- at the points of interpolation.

1. *Construction of the curve $y=Q(x)$, $Q\in\pi_{n-1}$, that is tangent alternatingly to C_+ and C_- at the points x_k.* The construction is both explicit and simple: Expanding the binomial

$$h(x)=(x+i)^n=P(x)+iQ(x),$$

where

$$P(x)=x^n-\binom{n}{2}x^{n-2}+\cdots,\qquad Q(x)=\binom{n}{1}x^{n-1}-\binom{n}{3}x^{n-3}+\cdots,$$

we have the identity

$$(g(x))^2=(x^2+1)^n=h(x)\overline{h(x)}=(P(x))^2+(Q(x))^2.$$

This implies that

$$-g(x)\leq Q(x)\leq g(x)\quad\text{for all real } x. \tag{3}$$

In polar form we have $x+i=(x^2+1)^{1/2}e^{i\theta}(0\leq\theta\leq\pi)$, with $x=1/\tan\theta$. Letting θ *increase from* 0 *to* π, we see that x *decreases from* $+\infty$ *to* $-\infty$. From $h(x)=(x+i)^n=(x^2+1)^{n/2}e^{in\theta}$ we see that the curve $\Gamma : z=P(x)+iQ(x)$ of the complex plane of $z=u+iv$, will spiral around the origin from the value of the argument $n\theta=0$ (hence $x=+\infty$) to $n\theta=n\pi$ (for $x=-\infty$). It follows that Γ will cut the v-axis in precisely n points for the values θ_k such that $n\theta_k=(2k-1)\pi/2$, hence for $\theta_k=(2k-1)\pi/(2n)$, for $k=1,\dots,n$. The corresponding values of x are

$$x_k=\left[\tan\frac{2k-1}{2n}\pi\right]^{-1},\qquad k=1,\dots,n.$$

From the spiraling motion we have for these values that

$$P(x_k)=0,\quad Q(x_k)=(-1)^{k-1}g(x_k),\quad\text{for } k=1,\dots,n.$$

This shows that $Q(x)$ is the desired interpolating polynomial.

2. *Proof that only $Q(x)$ gives the minimal value $S(x_1,\dots,x_n)=n$.* We assume $(\bar{x}_1,\dots,\bar{x}_n)\neq(x_1,\dots,x_n)$ and let $\bar{Q}(x)$ be the interpolating polynomial for the $\bar{x}_k$, so that $\bar{Q}(\bar{x}_k)=(-1)^{k-1}g(\bar{x}_k)$. From (3) we find that

$$(-1)^{k-1}(\bar{Q}(\bar{x}_k)-Q(\bar{x}_k))=g(\bar{x}_k)+(-1)^kQ(\bar{x}_k)\geq 0\quad\text{for } k=1,\dots,n,$$

while *not all these quantities vanish*. But then the divided difference of the polynomial $\bar{Q}-Q$ at the points $\bar{x}_1,\dots,\bar{x}_n$ is positive. This shows that *the difference* $\bar{Q}(x)-Q(x)=(\bar{S}-S)x^{n-1}+\cdots$ *has a positive highest coefficient* and so $S(\bar{x}_1,\dots,\bar{x}_n)>S(x_1,\dots,x_n)=n$.

Also solved by Allen Stenger and the proposer.

A Probabilistic Inequality

E 2888 [1981, 349]. *Proposed by T. F. Mori and G. J. Szekely, Lorand Eötvös University, Hungary.*

Let $A_1, A_2,\dots,A_n$ be an arbitrary sequence of events in a probability space. Prove that

$$\prod_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\frac{P(A_iA_j)}{P(A_i)P(A_j)}\geqslant 1$$

always holds if the probabilities $P(A_1)$ are positive.

Solution by the proposers. We prove the following generalization:

If $X_1, X_2,\dots,X_n$ are arbitrary nonnegative random variables with positive means and variances, then

$$\prod_{i=1}^{n}E(X_i\bar{X})\geqslant\prod_{i=1}^{n}E^2(X_i)$$

holds, where $\bar{X}=(1/n)(X_1+X_2+\cdots+X_n)$.

Replacing X_i by $I(A_i)/P(A_i)$ (where $I(A_i)$ denotes the indicator function of the event A_i, i.e., $I(A_i)=1$ on the event A_i and 0 otherwise) we obtain the inequality of the problem.

To prove the generalization we make the following self-explanatory steps:

$$1=\frac{1}{n}E\left(\sum_{i=1}^{n}\frac{X_i\bar{X}}{E(X_i\bar{X})}\right)=\frac{1}{n^2}E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{X_iX_j}{E(X_i\bar{X})}\right)$$

$$=\frac{1}{n^2}E\left[\left(\sum_{i=1}^{n}\frac{X_i}{\sqrt{E(X_i\bar{X})}}\right)^2+\sum_{1\leqslant i<j\leqslant n}X_iX_j\left(\frac{1}{\sqrt{E(X_i\bar{X})}}-\frac{1}{\sqrt{E(X_j\bar{X})}}\right)^2\right]$$

$$\geqslant\frac{1}{n^2}E\left(\sum_{i=1}^{n}E\sum_{i=1}^{n}\frac{X_i}{\sqrt{E(X_i\bar{X})}}\right)^2\geqslant\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E(X_i)}{\sqrt{E(X_i\bar{X})}}\right)^2\geqslant\left(\prod_{i=1}^{n}\frac{E^2(X_i)}{E(X_i\bar{X})}\right)^{1/n}.$$

Remark. The special case when the events A_i are exchangeable was announced to have been solved in the Institute of Mathematical Statistics Bulletin 9, 3, 1980 Apr., Masashi Okamoto: *On the Laslett Conjecture* (abstract).

Sum of Projections is Equal to Semiperimeter of an Equiangular Polygon

E 2889 [1981, 349]. *Proposed by I. J. Good, Virginia Polytechnic Institute.*

Let P be an arbitrary point in the plane of a regular polygon $A_1, A_2\dots A_n$. Let the foot of the perpendicular from P on line A_iA_{i+1} be Q_i (where A_{n+1} means A_1). Let x_i be $\pm$ length A_iQ_i: positive if Q_i, A_{i+1} are on the same side of A_i; negative otherwise. Prove that Σx_i is equal to half the perimeter of the polygon.

Solution by Anders Bager, Hjørring, Denmark. Let $|A_iA_{i+1}| = 1$. Then clearly $x_i = (\overrightarrow{A_iP}) \cdot (\overrightarrow{A_iA_{i+1}})$. Let O be the center. Then $\overrightarrow{A_iP} = \overrightarrow{A_iO} + \overrightarrow{OP}$, so that $\Sigma x_i = \Sigma(\overrightarrow{A_iO}) \cdot (\overrightarrow{A_iA_{i+1}}) + \overrightarrow{OP} \cdot \Sigma\overrightarrow{A_iA_{i+1}}$. But the last sum is 0. This reduces the assertion to the case in which $P = O$. □

Also solved by K. L. Bernstein, J. C. Binz (Switzerland), P. S. Bruckman, B. Cheng & D. T. Hung (students), J. Dou (Spain), H. Eves, J. Grave & J. Papes (Spain), T. Hermann (Hungary), V. Hernandez (Spain), H. Kappus (Switzerland), L. R. King, L. Kuipers (Switzerland), I. G. Macdonald (UK), W. A. Newcomb, R. W. K. Odoni (UK), I. Paasche (Germany), M. R. Railkar (India), D. Rawsthorne, B. L. R. Shawyer (Canada), R. Spaulding, H. R. van der Vaart, and the proposer.

Macdonald noted that the proof applies even when the n-gon is an n-pointed star. Spaulding cited L. H. Miller, *College Geometry*, and gave another generalization. Paasche used an averaging method. Eves noted that Σx_i^2 is also unchanged when P is moved to O. Dou mentioned E 2885, and noted that the n-gon need only have a rotational symmetry (by $2\pi/g$, g integer), or be equiangular.

Map [0, 1] ⇔ [0, 1] with $f(x - f(x) + x) = x$

E 2893 [1981, 444]. *Proposed by Ko-Wei Lih, Institute of Mathematics, Academia Sinica, Taipei.*

Determine all homeomorphisms f from $[0, 1]$ to $[0, 1]$ that are solutions of the functional equation $f(2x - f(x)) = x$ for all x in $[0, 1]$.

I. *Solution by Fred Richman, New Mexico State University.* Continuity of f and f^{-1} is irrelevant. The only solution is $f(x) \equiv x$. By hypothesis, $f^{-1}(x) = 2x - f(x)$, so that $f(x) - x = x - f^{-1}(x)$. Now starting with any x_0 in $[0, 1]$, define x_n inductively by $x_n = f(x_{n-1})$. Setting $x = x_{n-1}$ in the above equation, we obtain $x_n - x_{n-1} = x_{n-1} - x_{n-2}$. It follows that $x_n - x_{n-1} = x_1 - x_0$ for all n, and hence $x_n - x_0 = n(x_1 - x_0)$. Since $|x_n - x_0| \leq 1$, we conclude that $|x_1 - x_0| < 1/n$ for all n, and therefore $|x_1 - x_0| = 0$. Thus $f(x_0) = x_1 = x_0$.

II. *Solution by I. M. Isaacs, University of Wisconsin, Madison.* In fact, we need not even assume that f is one-to-one (or equivalently, that f^{-1} exists). Write $d(x) = x - f(x)$ for x in $[0, 1]$. Suppose that $d(x_0) = a$. Then $a + x_0 = 2x_0 - f(x_0)$ and so, by hypothesis, this number lies in $[0, 1]$ and $f(a + x_0) = x_0$. Thus $d(a + x_0) = a = d(x_0)$. Therefore, the complete inverse image $d^{-1}(\{a\})$ is a subset of $[0, 1]$ which is mapped into itself by the addition of a. It follows that $a = 0$ and $x - f(x) \equiv 0$. (Essentially the same solution was given by Bernstein, Crofts, Elridge, Gustafson, Mauldon, Myerson, and Wu.)

Bencze generalized the hypothesis to $f((k + 1)x - kf(x)) = x$ for real k with $k \geq 1$ or $k < -1$.

III. *Solution by D. Laugwitz, Technische Hochschule, Darmstadt, Germany.* When suitably interpreted, the proposition holds when an arbitrary group G replaces the additive group of the real numbers. The condition that $f(2x - f(x)) = x$ is replaced by the equivalent multiplicatively written condition $f(xf(x)^{-1}x) = x$.

THEOREM. *Let D be a subset of G and suppose f is a mapping on D such that $f(xf(x)^{-1}x) = x$ for all $x \in D$. If D contains no left coset Hx of a nonidentity subsemigroup H, then $f(x) \equiv x$ on D.*

The proof parallels that in Solution II with $d(x) = xf(x)^{-1}$. If $d(x_0) = a$ then $d(ax_0) = a$ and so $Hx_0 \subseteq D$ where $H = \{a^n \mid n \geq 0\}$. Thus $a = 1$ and $d(x) = 1$ for all x.

Also solved by U. Abel (student, Giessen), V. Anantharam, M. D. Ašić (U.K.), M. Bencze (Romania), K. Bernstein, I. Bivens, J. Bobek (student, Czechoslovakia), D. C. Buchthal, J. Browkin (Poland), R. Cabane (France), F. S. Cater, Chico Problem Group, G. Crofts, M. W. Ecker, P. Eenigenburg, S. Elridge (U.K.), A. Epstein, G. Fisher, E. Grosswald, N. Franceschine, W. H. Gustafson, G. C. Harrison, D. Hart, V. Hernandez (Spain), G. A. Heuer, K. J. Heuvers, D. T. Hùng & B. Cheng (students), E. D. Huthnance, E. Johnston, W. T. M. Kars (Netherlands), R. Kowalski, M.-L. Kung, D. Laugwitz (Germany), J. Levy, J. S. Lew, J. G. Mauldon, G. J. Michaelides, F. B. Miles, M. D. Meyerson, W. A. Newcomb, H. W. Oliver, V. Pambuccian (Rumania), D. A. Rawsthorne, D. Ross, M. B.

Ruskai, P. J. Ryan, I. A. Sakmar (Canada), B. Schaaf (Netherlands), M. St. Vincent & R. St. Vincent, I. J. Schoenberg, F. B. Strauss, R. A. Struble, P. Y. Wu (Taiwan), W. R. Utz, L. E. Ward Jr., D. M. Wells, J. Wichmann (Saudi Arabia), J. B. Wilker (Canada), M. F. Wyneken, F.-S. Yen (Taiwan), K. L. Yocom, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, The University of Western Ontario, London, Ontario, Canada N6A 5B9, by March 31, 1983. The solver's full post-office address should be on each sheet.

6403. *Proposed by J. L. Brenner, Palo Alto, CA.*

Let $0 < r, s$. For positive reals, $A, B, A \neq B$ define

$$M_1 = \left[\frac{1}{2}(A^rB^s + A^sB^r)\right]^{\frac{1}{r+s}}$$

$$M_2 = \left[\frac{1}{2}(A^{rs} + B^{rs})\right]^{\frac{1}{rs}}$$

Since $M_1 < M_2$ as $B \to \infty$, either $M_1 < M_2$ for all A, B or else M_1, M_2 are incomparable. Which is the case?

6404. *Proposed by Daniel Shanks, University of Maryland.*

The imaginary quadratic field $\mathbb{Q}(\sqrt{-2 \cdot 5 \cdot 2347 \cdot 10513})$ has

$$L(1, \chi) = 0.5008000001\ldots,$$

where, in

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) \cdot n^{-s},$$

$\chi(n) = (d/n)$ is the Kronecker symbol and d is the discriminant. Show that one can find imaginary quadratic fields where the decimal value of $L(1, \chi)$ contains a string of zeros that is as long as one wishes.

6405. *Proposed by Dragoljub Milosevic, Pranjani, Yugoslavia.*

Let $T_n(x)$ be the nth Chebyshev polynomial ($n > 1$). Prove or disprove

$$\int_{-1}^{1} \frac{x}{\sqrt{1-x^2}} T_n^3(x)\, dx < \pi\sqrt{2}/8$$

6406. *Proposed by J. Michael Steele, University of Chicago.*

Show that if $\sum_{n=0}^{\infty} a_n e^{int} = f(t)$ and $f(t) \in L^1[0, 2\pi]$, then there exist integers n_k such that

$$n_k \to \infty,\ n_{k+1}/n_k \to 1, \text{ and } \sum_{k=1}^{\infty} |a_{n_k}| < \infty.$$

6407. *Proposed by L. Van Hamme, Free University of Brussels, Belgium.*

Define $\begin{bmatrix} n \\ k \end{bmatrix}$ by means of the relation

$$\begin{bmatrix} n \\ k \end{bmatrix} = F_{n,k}/F_{k,k},\ F_{n,k} = (q^n - 1)(q^{n-1} - 1) \cdots (q^{n-k+1} - 1),$$

so that $\begin{bmatrix} n \\ k \end{bmatrix}$ is the so-called Gaussian polynomial. Prove the identity

$$\sum_{k=1}^{n} \frac{q^k}{1-q^k} = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{1-q^k} q^{1/2k(k+1)} \begin{bmatrix} n \\ k \end{bmatrix}.$$

6408. *Proposed by Alexander Kovaćec, University of Vienna, Austria.*

Let $T := \{z \in \mathbb{C}: |z| = 1\}$. Let λ denote Lebesgue measure on T, normalized so that $\lambda(T) = 1$. For a given positive integer n let A be a subset of T with $\lambda(A) > 1 - 1/n$. Now suppose B is a set in T containing exactly n points. Show that there exists a rotation of B which carries it into A, i.e., there exists $c \in T$ such that $cB \subset A$.

SOLUTIONS OF ADVANCED PROBLEMS

Determining Heavy and Light Balls by Weighings

6224 [1978, 600; 1980, 828]. *Proposed by David P. Robbins, Hamilton College.*

Suppose we are given N balls which are indistinguishable except that some are heavy and some are light (the heavy balls are alike in weight, as are the light balls). Using a pan balance what is the minimum number of weighings in which it is always possible

(a) to identify one heavy and one light ball?

(b) to determine the number of heavy and light balls?

Solution by Kiyoshi Takizawa, Tokyo Gakugei University, Japan. Zane C. Motteler's partial solution to this problem can be improved. We consider the following weighing process.

Let A be the set of n balls ($n \geqslant 2$). *We divide A into three parts B, C, and D, where B and C have $[n/2] = n\alpha$ balls and $D = 0$ or 1 according as n is even or odd. Set $n\beta = -[-n/2]$. α and β are mappings of the set of integers into itself as follows:

$$n\alpha = \left[\frac{n}{2}\right], \quad \text{the greatest integer not greater than } \frac{n}{2};$$

$$n\beta = -\left[-\frac{n}{2}\right], \quad \text{the smallest integer not smaller than } \frac{n}{2}.$$

Weigh B with C.

(1) If they balance, then we regard $B \cup D$ as A and repeat*.

(2) If they do not balance, then **we divide B, C into two parts, B_1, B_2 and C_1, C_2, respectively, where B_1 and C_1 have $(n\alpha)\alpha$ balls, and B_2 and C_2 have $(n\alpha)\beta$ balls.

Weigh B_1 with C_1.

(2a) If they balance, then we regard B_2 and C_2 as B and C respectively and repeat**.

(2b) If they do not balance, then we regard B_1 and C_1 as B and C respectively and repeat **.

Using the process above, we can identify one heavy and one light ball in at most k weighings if $2^{k-1} < n \leqslant 2^k$.

Closed Graph Theorem

6255 [1979, 132; 1980, 679]. *Proposed by Adam Riese, Wright State University, Dayton, OH.*

Let $f: R \to R$ be a function whose graph, considered as a subset of R^2, is both closed and connected. Prove f is continuous. What can be said when $f: R^m \to R^n$?

F. S. Cater has pointed out that the published solution cannot be extended to $f: R^m \to R^n$ if

$m > 1$, since f may be discontinuous. Cater and Tripp agree that the solution covers the case $m = 1$ and, in particular, the case $m = n = 1$.

Product of p^mth powers in A_n

6315 [1980, 676]. *Proposed by Jan Mycielski, University of Colorado, and the editors.*

Let $l = p^n$ be a prime power. Find the smallest $k = k(n, l)$ such that every element in the alternating group A_n can be expressed as a product of k lth powers.

Solution by Gade Moran, University of Haifa, Israel. The value of k is almost always 1 or 2. If $(|A_n|, p) = 1$, i.e., if $n < p$, then every element of A_n has order prime to l, so is an lth power of the element itself. Hence $k = 1$.

If $n \geqslant 10$, $p < n$, there is a number t between $[3n/4]$ and n (inclusive) that is prime to p. By the theorem of Bertram-Brenner, every element of A_n is expressible as the product of two t-cycles, and hence of two lth powers. Thus $k = 2$. (There exists a cycle in A_n that is not an lth power.)

The cases $n < 10$ have to be checked one at a time. The result is $k(1, l) = k(2, l) = 1$; $k(3, l) = 1$ if $(l, 3) = 1$; $k(3, 3^m) = \infty$; $k(4, 2^m) = 2$, $k(4, 3^m) = \infty$; $k(n, l) = 2$ for $5 \leqslant n \leqslant 10$. The reference for these results, namely that the square of the type $2^1(n-2)^1$ covers A_n, is Xu-Brenner.

References

1. E. Bertram, Even permutations as a product of conjugate cycles, J. Comb. Theory Ser. A, 12 (1972) 368–369.
2. J. L. Brenner, Covering theorems for nonabelian simple groups, 9, Ars Combin., 4 (1977) 151–176.
3. C. -H. Xu, The commutators of the alternating group, Sci. Sinica, 14 (1965) 339–342.

Comment by the editors. Similar results hold when l is not a prime power. All that is needed is that a number exist in the range $([3n/4], n)$ that is odd and prime to l, or that a number exist in the range $([3n/4], n-2)$ that is even and prime to l. For example, if $l = 30$, then $k(n, l) = 2$ whenever n exceeds 20. See the references cited above.

Partition of S into n triples

6316 [1980, 758]. *Proposed by David Winter, University of Michigan.*

Let S be a set of $3n$ points in $\mathbb{R}^3$, no four of which are coplanar. Suppose that $S = R \cup Y \cup G$, where each of R, Y, G has n points. Is it possible to partition S into triples $\{r_i, y_i, g_i\}$, $1 \leqslant i \leqslant n$, where each r_i is in R, each y_i is in Y, and each g_i is in G, in such a way that the n triangles $T_i = \operatorname{conv}\{r_i, y_i, g_i\}$ are pairwise disjoint?

Solution by Karl W. Heuer, Moorhead, Minnesota. It is possible to find such triples, as will now be proved by induction.

When $n = 1$ the solution is trivial. When $n > 1$, construct about each point a ball of radius ε, where ε is chosen so that no plane intersects more than 3 of the balls. By the three dimensional case of the "Ham Sandwich" Theorem [**1**], [**2**], [**3**], there is a plane which simultaneously bisects the volume of the three sets. This plane may intersect some of the balls. If n is even, the plane may be adjusted if necessary so that exactly $n/2$ of the points of each set are on each side. If n is odd, the plane must pass through one point of each set; join these with a triangle, and there are $(n-1)/2$ remaining points of each set on each side. In either case these two sets of $3[n/2]$ points may be independently partitioned, using the induction hypothesis. The resulting partition of S completes the induction. This argument generalizes to R^d, $d > 3$.

Essentially the same proof was given by Richard Goldstein, SUNY Albany.

References

1. D. W. Blackett, Elementary Topology: A Combinatorial and Algebraic Approach, Academic Press, New York (1967) 119-120.

2. A. H. Stone and J. W. Tukey, Generalized "Sandwich" Theorems, Duke Math. J., 9 (1942) 356–359.

3. H. Steinhaus, Sur la division des ensembles de l'espace par les plans et des ensembles plans par les cercles, Fund. Math., 33 (1945) 245–263.

Also solved by the proposer.

An Undecidable Question on Sequences of Integers

6333 [1981, 150]. *Proposed by M. J. Pelling, London, UK.*

Let N_* denote the set of infinite integer sequences $n_* = (n_1, n_2, \dots)$, all components being nonnegative, partially ordered according to: $m_* \leqslant n_*$ if $\forall_i[m_i \leqslant n_i]$. Show that the following question is undecidable. "Does N_* contain an uncountable subset X_* such that, for every member k_* of N_* the set

$$\{n_* | n_* \in X_* \quad \text{and} \quad n_* \leqslant k_*\}$$

is countable?"

Solution by Brian M. Scott, Cleveland State University. For $f, g \in {}^\omega\omega$ write $f < {}^*g$ iff there is an $m \in \omega$ such that $f(n) < g(n)$ whenever $n \geqslant m$. Now consider the following assertions:

(A) There is an uncountable $X \subseteq {}^\omega\omega$ such that for each $f \in {}^\omega\omega$, $\{x \in X: x \leqslant f\}$ is countable.
(B) For any $X \subseteq {}^\omega\omega$ of cardinality ω_1 there is an $f \in {}^\omega\omega$ such that $x < {}^*f$ for each $x \in X$.
1. (B) implies $\neg$(A).

Proof. Fix $X \subseteq {}^\omega\omega$ of cardinality ω_1, and let f be as in (B). (Plainly in (A) we may assume that $|X| = \omega_1$.) For each $x \in X$ there is an $m_x \in \omega$ such that $f(n) > x(n)$ whenever $n \geqslant m_x$. But then there are $n \in \omega$, a finite function $p: m \to \omega$, and an uncountable $Y \subseteq X$ such that for each $x \in Y$, $m_x = m$, and $x(n) = p$ for all $n < m$. Define $h: \omega \to \omega$ as follows:

$$h(n) = \begin{cases} p(n) + 1, & \text{if } n < m \\ f(n), & \text{if } n \geqslant m. \end{cases}$$

Then $x(n) < h(n)$ for all $x \in Y$ and $n \in \omega$; i.e., $Y \subseteq \{x \in X: x \leqslant h\}$, which is therefore uncountable.

2. $\neg$(B) *implies* (A).

Proof. Let $X = \{x_\alpha: \alpha < \omega_1\}$ be a counterexample to (B). Define $Y = \{y_\alpha: \alpha < \omega_1\} \subseteq {}^\omega\omega$ by recursion on α as follows. First, $y_0 = x_0$. If, now, $\alpha < \omega_1$, and y_β has been defined for each $\beta < \alpha$, let $Z = \{y_\beta: \beta < \alpha\} \cup \{x_\alpha\}$. Z is countable, so reindex it as $\{z_n: n \in \omega\}$. Now define y_α as follows:

$$y_\alpha(n) = 1 + \max\{z_i(n): i \leqslant n\}.$$

Then $y_\alpha(n) > z_i(n)$ for every $n \geqslant i$, so $z_i < {}^*y_\alpha$ for each $i \in \omega$. This completes the construction of Y.

Observe that Y has the following properties: $y_\alpha < {}^*y_\beta$ whenever $\alpha < \beta < \omega_1$, and $x_\alpha < {}^*y_\alpha$ for each $\alpha < \omega_1$. Clearly, then, Y is also a counterexample to (B). Now let $f \in {}^\omega\omega$ be arbitrary. There is an $\alpha < \omega_1$ such that $y_\alpha \not< {}^*f$, and (by passing to $y_{\alpha+1}$ if necessary) we may even assume that $y_\alpha \not\leqslant {}^*f$. That is, $\{n \in \omega: y_\alpha(n) > f(n)\}$ is infinite. It immediately follows that $y_\beta \not\leqslant {}^*f$ whenever $\alpha \leqslant \beta < \omega_1$, so that $\{y \in Y: y \leqslant f\} \subseteq \{y_\beta: \beta < \alpha\}$, which is countable, and Y satisfies (A).

In short, (A) is equivalent to $\neg$(B).

Now, $\neg$(B) is an easy and well-known consequence of the Continuum Hypothesis (CH): take X to be ${}^\omega\omega$ itself. Thus (A) is consistent (with ZFC). And (B) is a fairly well-known consequence of MA + $\neg$CH (MA = Martin's Axiom), so (A) is independent of ZFC as well. (For this connection see **[2]**, **[3]**, **[4]**, and for a general discussion of Martin's Axiom, see **[10]**.)

Scott goes on to note, as does Koepke, that (A) is not equivalent to CH, and refers to Hechler **[1]**.

Fred Galvin, University of Kansas, notes that the assertion (B) was first considered by Rothberger, under the name $B(\aleph_1)$, and that Rothberger [6], [7], [8], [9] found a variety of interesting equivalents of (B). He notes also that Pelling's statement implies $\neg B(\aleph_1)$, and that the implications CH $\Rightarrow \neg B(\aleph_1)$ and MA$(\aleph_1) \Rightarrow B(\aleph_1)$ are well known (see [5]).

Also solved by Dryden Cope, P. Erdös (Hungary), R. Koepke (Germany), and Simon Thomas (England).

References

1. S. H. Hechler, On the existence of certain cofinal subsets of ${}^{\omega}\omega$, Proc. Symp. Pure Math. AMS., 13 (1979) part II, 155–173.
2. T. J. Jech, Set Theory, Academic Press, New York, 1978, p. 261.
3. K. Kunen, Set Theory, North-Holland, 1980, p. 87, exercise 8.
4. Azriel Lévy, Basic Set Theory, Springer-Verlag, Berlin and New York, 1979, p. 282.
5. D. A. Martin and R. M. Solovay, Internal Cohen extensions, Ann. Math. Logic, 2 (1970) 143–178.
6. F. Rothberger, Sur un ensemble toujours de première catégorie qui est dépourvu de la propriété C, Fund. Math., 32 (1939) 294–300.
7. ______, Sur les familles indénombrables de suites de nombres naturels et les problèmes concernant la propriété C, Proc. Cambridge Philos. Soc., 37 (1941) 109–126.
8. ______, On some problems of Hausdorff and of Sierpiński, Fund. Math. 35 (1948) 29–46.
9. ______, On the property C and a problem of Hausdorff, Canad. J. Math., 4 (1952) 111–116.
10. J. R. Schoenfield, Martin's axiom, this MONTHLY, 82 (1975) 610–617.

Dense Orbits of Ergodic Transformations

6337 [1981, 213]. *Proposed by R. Daniel Mauldin, North Texas State University, and Jan Mycielski, University of Colorado.*

Prove that if h is a measure-preserving ergodic homeomorphism of the unit square S onto itself, then there exists a set $D \subset S$ such that D is of measure 1, $S - D$ is meager and for every $p \in D$ the sequence $p, h(p), hh(p), \dots$ is dense in S.

Solution by Adam Fieldsteel, Wesleyan University, Middletown, CT. Let $U_1, U_2, \dots$ be a basis of open sets for S. For each n the set $A_n = \cup_{i=1}^{\infty} h^{-i}U_n$ is open and $A_n \subseteq hA_n$. Hence, by the ergodicity of h, A_n is also of measure 1 and dense in S. Since

$$D = \bigcap_{n=1}^{\infty} A_n = \{p : \forall n \exists i [h^i(p) \in U_n]\}$$

it follows that D has all the desired properties.

Compare this with the Poincaré recurrence theorem and related results in J. Oxtoby, Measure and Category, Springer-Verlag, New York, 1970.

Also solved by Dinh Thê Hùng & Benny Cheng (students), Vincent C. Peck, James B. Robertson, Abraham Smuckler (Israel), and the proposers.

MISCELLANEA

87. This winter I am giving two courses of lectures to three students, one of whom is only moderately prepared, another less than moderately and the third lacks both preparation and ability. Such are the burdens of a mathematical profession.

—Gauss to Bessell, 1810. (Quoted in R. E. Mortiz, *Memorabilia Mathematica*, 1914, p. 158. Suggested by R. Askey.)

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Advanced Calculus. An Introduction to Modern Analysis. By William L. Voxman and Roy H. Goetschel, Jr. Marcel Dekker, New York, 1981. xii + 678 pp.

Foundations of Mathematical Analysis. By Richard Johnsonbaugh and W. E. Pfaffenberger. Marcel Dekker, New York, 1981. viii + 428 pp.

WILFRED KAPLAN
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109

These textbooks are intended to bridge the gap between elementary calculus, a highly standardized subject, and higher analysis and its applications, a subject of enormous extent, and one in a continual state of flux.

Filling the gap presents writers a number of difficult problems of selection of topics, depth of treatment, choice of applications, mathematical rigor, providing suitable exercises.

In many ways, the easiest way to resolve many of the difficulties is to view analysis as a rigid logical structure and, accordingly, to state one's axioms and proceed to develop a significant part of the subject in strict definition-theorem-proof format. Both books under review pursue this approach to some extent.

Unfortunately, this type of presentation has a number of serious drawbacks. A complete logical development is necessarily long, tedious, repetitive. (The completeness may of course be illusory, in any case, because of foundation questions.) It is often a poor way to present the true "flavor" of a topic. It excludes from discussion many important areas, such as those relying on the Jordan theorem for the plane. It leads one to clumsy formulations, often abandoned by the writers themselves after a stretch, in favor of the traditional, "less rigorous" terminology. An example, occurring in both books, is the definition of an infinite series as an ordered pair of sequences $(\{a_n\}, \{s_n\})$, where $s_n = a_1 + \cdots + a_n$.

Accordingly, writers on advanced calculus are forced to compromise on the rigorous, complete development. In the book by Johnsonbaugh and Pfaffenberger, there are recognized omissions in the treatment of the real number system, with a reference to the classical book of Landau; the remaining subjects are treated in essentially rigorous fashion. However, the authors have chosen a set of topics which are amenable to a strict development: real sequences and series, metric spaces, differential and integral calculus (including the Riemann-Stieltjes integral) for real functions, Fourier series, normed linear spaces, the Lebesgue integral on an abstract measure space. The geometric aspects of analysis are ignored. Line integrals, multiple integrals, Green's theorem and related topics are not mentioned. Physical or other applications are not covered.

The text by Voxman and Goetschel places far less emphasis on rigor, and gives correspondingly more weight to motivation and intuition. In fact physical problems such as the vibrating string are used to motivate the treatment of series of functions and other mathematical topics. Line integrals and Green's theorem are treated. Furthermore, most of the topics covered by Johnsonbaugh and Pfaffenberger are included, generally in less depth. The formula for change of variables in multiple integrals is treated only briefly in a problem; the other text ignores multiple integrals.

The result of the selection processes in both cases is the omission of a number of topics of importance. Whether students are being educated for graduate research in mathematics, for becoming teachers of mathematics, or for work in other fields, it seems unfortunate to exclude such heavily used tools as Stokes' theorem. In fact, I would argue that the "fundamental theorem of calculus" in all its manifestations (which include Stokes' theorem and the change of variables formula) is at the heart of analysis and its applications. In fact, it has profound links to geometry

and topology (see for example *Geometric Integration Theory* by Hassler Whitney, Princeton University Press, 1957). To ignore or treat only superficially this core topic is to deny the student an essential ingredient of mathematical education. To downgrade or suppress the many ties between geometry and analysis is a similar disservice to the student. A justification for the omission or downgrading which might be offered is that geometric analysis is difficult to treat rigorously, and that it is preferable to choose topics such as elementary metric space theory which can be developed easily in axiomatic form. In my judgment, this justification is insufficient and it is preferable to yield some on rigor in order to convey some ideas of inestimable value.

The long rigorous developments chosen also lead the student to advanced areas, far above the level of elementary calculus, and one can wonder about the wisdom of trying to reach these areas so early in the student's education. Examples of such advanced topics are the convergence theorems for Lebesgue integrals and the Riesz representation theorem. The accompanying illustrative examples (when provided) or exercises are often very elementary or difficult theoretical questions.

Thus the middle ground of analysis is a difficult one for textbook writing. The authors of the books reviewed here have made strenuous efforts to provide clear logical presentations and have generally succeeded in this. The principal unresolved questions are the choice and level of topics and the desirability of sacrificing some rigor in order to reach some important results which belong to the general education of mathematicians, pure and applied.

An Introduction to Data Analysis. By Bruce D. Bowen and Herbert F. Weisberg. W. H. Freeman, San Francisco, 1980. x + 213 pp.

DAVID S. MOORE
Department of Statistics, Purdue University, West Lafayette, IN 47907

The first course in statistical methods is a staple of the campus diet, often served up with varying degrees of sophistication by several different academic units. Mathematics departments have tended to prefer a gourmet version, well seasoned with Venn diagrams and combinatorics, and perhaps even requiring a palate accustomed to calculus. Other providers, particularly social science departments, have stressed methods and attention to the needs of science over theory. While typical expressions of both types of course have much to criticize, the psychologists and sociologists are clearly right in their basic choice. Statistics is a methodological science, not a branch of mathematics. The tools of mathematics are of course essential to statisticians, and mathematical research on problems arising in statistics is a valuable and active field (my own field, in fact). But attitudes grounded in mathematical research have no place in an introduction to statistical methods intended for a wide audience. On the other hand, social and behavioral scientists have often been reluctant to ask their students to master high school mathematics before studying statistics, and have thus condemned themselves to an eternity of teaching algebraic illiterates how to substitute numbers into formulas.

The deficiencies of both theorem-proof and cookbook approaches, and the pressure of new developments, appear to be leading to a welcome reassessment of the first statistics course, and to a rough consensus that the essential objectives are to teach students the art of examining data and certain modes of reasoning that are common to the many specific methods of inference. A number of texts reflect this approach, from my own book for humanities students [5] and the work of Freedman, Pisani and Purves [3] (which stops somewhat short of covering the usual material of the first methods course) to the more advanced methodological expositions of Box, Hunter, and Hunter [1] and Mosteller and Tukey [6]. A mathematician wishing to understand the side of statistics that is not mathematics could profitably read these books in place of the next four light

novels in his queue. They are all good.

Several recent changes in statistics will affect the first methods course. The most obvious change (in almost any discipline) is the universality of computers. Inexpensive computing allows processing of greater quantities of data, and also data of greater complexity (many variables, missing data, unbalanced designs). Perhaps more significantly, computers allow the use of new statistical methods involving "previously unthinkable calculations." That phrase comes from Efron [2], who argues for reconsideration of standard methods that were developed under the constraint of slow and expensive calculation. Thus the specific methods taught in a first course are no longer those most likely to be used in practice, now or in the future. Moreover, the arithmetic of the methods that will be employed will certainly be automated. These facts argue, in my opinion, for increased attention to the difficult questions and ideas that undergird both simple and complex methods. What specific methods are used to convey a broader understanding is not a critical issue.

A second, and more controversial, development is the rise of "exploratory data analysis" or EDA. John Tukey, the major innovator in the field, has described EDA as a flexible attitude combined with lots of graph paper. EDA is the art of letting the data speak without imposing a prior structure, using the unique ability of the human eye to detect the unexpected. Traditional inference, which seeks answers to clearly posed questions based on a model that is (we hope) grounded in a planned design for data collection, is called "confirmatory" in contrast. Hartwig and Dearing [4] give a particularly clear exposition of the EDA perspective. Many of these ideas, along with Tukey's innovations in descriptive statistics, are appearing in introductory courses and texts.

But EDA is easily criticized as describing phenomena without offering any underlying explanation. This backlash brings with it a renewed appreciation for an essential aspect of statistics that has been largely ignored in basic courses: The design of data collection. Statistics (like all Gaul) can be divided into three parts: collecting data, organizing and describing data, and inference from data. EDA has called attention to the second (directly) and the first (indirectly) of these areas. A balanced introduction to statistics must stress the role of properly designed data collection and (admitting that good data are rare) the necessity of skillful data analysis, as well as formal inference. That a properly randomized design tends to validate common models is a mathematical fact, but issues such as the relative roles of exploratory and confirmatory analysis concern scientific method rather than mathematics.

A third recent development of considerable importance is the study of robustness, the degree to which methods of inference remain valid when their assumptions are not met. Since in practice the model is never quite correct, some degree of robustness is essential to a usable procedure. Both the idea of robustness and some facts about the robustness of standard methods should appear in a first course. The most commonly taught procedures share the assumption that the underlying population has a normal distribution, but not all normal-theory procedures are equally robust. The two-sample t-test for means, for example, is quite robust against departures from normality, at least when the sample sizes are similar. But the two-sample F-test for variances is so sensitive to lack of normality that it cannot be recommended for use. Dramatic graphical comparisons of these two common tests can be found in Pearson and Please [7]. It is now negligent to present these tests on the same plane, as simply the two-sample tests for means and variances, respectively. Their equal mathematical status as likelihood ratio tests in the normal distribution case has little bearing on actual utility.

These varied developments may appear to overwhelm an introductory course; they certainly do allow wide variations in content and emphasis. It is important, in my opinion, that instructors not be swept away in a newly fashionable direction. Researchers in robustness, for example, offer a variety of more trustworthy tools in addition to evaluations of standard tools. But there is much to be said for concentrating on a selection of older methods in a first course, since it is not yet clear which robust procedures will become standard and (more fundamentally) because classical methods force the issue of stating and assessing a model. Similarly, unalloyed EDA is not a

responsible introduction to statistics. Computers can also seduce the unwary, resulting in a "black box" approach in which methods for analyzing complex data are presented without adequate conceptual foundations. (I do not mean to attack pedagogical use of computers, only the immediate introduction of statistical methods so complex as to require computer arithmetic. Interactive computing facilities in particular seem to me an excellent medium for teaching statistical methods.)

I come then to recommending that students be exposed to certain basic ideas and skills and to a selection of common statistical methods, taught with emphasis on data and on the conceptual place of statistics in science. Data collection, the power of plotting, the habit of assessing the model from the data, and the logic of classical inference should be stressed. A student should come away from a first statistics course respecting experimentation as the preferred source of knowledge, knowing how to check normality by simple plots, and aware of such pitfalls in inference as the perils of aggregation and the multiple comparisons problem. Exactly because computations are automated, the student can very rapidly build a knowledge of additional techniques on the solid foundation provided.

The spirit that I have outlined here is spreading. The four books mentioned above all embody variations of it, though none is fully satisfactory for teaching undergraduates to use statistical methods. Bowen and Weisberg have somewhat similar goals. They emphasize that "data analysis" is broader than "statistics," which they identify with statistical inference. Their book contains many sensible comments on looking at data. Their exposition concentrates on traditional methods, and is clear and generally accurate. But the exercises are few and skimpy, and there is no adequate discussion of the meaning of common modes of inference such as tests and confidence intervals. I was also unhappy at the attempt to say something about a very wide variety of procedures whose actual mechanics are not described at all. The exposition in the latter part of the book is therefore very loose and general—six pages on multiple regression, four pages on factor analysis, two on multidimensional scaling, and so on. This style is justified in part by the stated goal of preparing students to read literature in which data are analyzed, rather than to work with data themselves. It also reflects the unfortunate reluctance of some disciplines to require an adequate quantitative background of students. The sample mean first appears on page 145 (of 213), and its arithmetic is painstakingly explained. *Introduction to Data Analysis* is therefore not a text of methodology. It is a thoughtful description for potential users, one that responds in part to the trends discussed above. But only in part: data collection, the power of plotting, and the logic of inference are slighted, and the role of models and robustness questions are ignored. To my taste, mathematically unprepared students would be better served by Freedman, et al., which gives a much deeper understanding of fundamental issues and hands-on experience in interpreting simple data at the sacrifice of broad coverage of fancier techniques.

The central theme of this review has been the insistence that a first statistics course should emphasize understanding data rather than presenting mathematics. Bowen and Weisberg are among the authors who share this emphasis. I hope that widespread recovery of the idea that statistics has its conceptual setting in the methodology of science will both improve the usefulness of the first course and relieve its reputation for deadening dullness.

References

1. G. E. P. Box, W. G. Hunter, and J. S. Hunter, Statistics for Experimenters, Wiley, New York, 1978.
2. B. Efron, Computers and the theory of statistics: thinking the unthinkable, SIAM Rev., 21 (1979) 460–480.
3. D. Freedman, R. Pisani, and R. Purves, Statistics, Norton, New York, 1978.
4. F. Hartwig and B. E. Dearing, Exploratory Data Analysis, Sage Publications, Beverly Hills, 1979.
5. D. S. Moore, Statistics: Concepts and Controversies, Freeman, San Francisco, 1979.
6. F. Mosteller and J. W. Tukey, Data Analysis and Regression, Addison-Wesley, Reading, MA, 1977.
7. E. S. Pearson and N. W. Please, Relation between the shape of the population distribution and the robustness of four simple test statistics, Biometrika, 62 (1975) 223–241.

Finite and Infinite Dimensional Linear Spaces. By Richard D. Järvinen. Marcel Dekker, New York, 1981, xiv + 168 pp.

HERON S. COLLINS
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803

In this pleasant little book, the author tries with some success to give a selected comparison of finite versus infinite dimensional linear spaces. He develops the program nicely, first having a theorem for linear spaces or linear topological spaces with finite Hamel basis, and then commenting on the infinite dimensional version. Numerous (for such a brief treatment) examples and problems are given as the book progresses which help in showing the contrasts and comparisons between the finite and infinite cases.

All of the above was pleasant enough to this reviewer and not at all unexpected, once I got accustomed to the idea. However, I found more interesting and less familiar the other main thread which Dr. Järvinen chose to weave in his monograph, viz., the generalizations and extensions of finite systems of linear algebraic equations to infinite systems to integral equations, and finally to a study of special types of linear operators (Volterra, Fredholm, Compact), and the inclusion of some theorems dealing with infinite matrices and infinite systems I'd never seen before. As a functional analyst I found this very nice, and the author's evident historical expertise and interest in things historical (for example, in the first chapter and a half one finds references to Kronecker, Paul Cohen, Cantor, Gödel, Hille, Poincaré, Hilbert, Pólya, Bôcher, and others, and comments about their connections with the mathematics under discussion) makes the work flow well.

A third aspect of this effort which I found a bit unexpected though welcome is the inclusion by the author of a number of "challenge" and "open" problems both in the theory of linear topological spaces and in general infinite dimensional topology. The latter reflects, I suspect, the strong influence of Dr. Järvinen's friend and colleague, T. Rassias, who wrote the foreword to the book and who is often referred to, especially in the last chapter, Chapter 4. Some of the challenge problems are indeed a challenge, and, in fact, at the end of *each* chapter there is a generous number of challenge problems as well as more pedestrian ones for the less experienced student. On the other hand, some of the "open" problems are (according to my local experts) no longer open, but that is neither surprising in fast-moving fields, nor need it be particularly disturbing to the reader.

It may well be that the purist students of linear or linear topological spaces, accustomed as they are to more traditional and existing texts or expositions, will question the need for a monograph such as this one. To be quite candid about it, I had some doubts myself at first glance as to its potential usefulness. After all the audience for books or even small monographs in mathematics is relatively small, and publishers of such books are rare and hard to come by. On the other hand, the primary concern should be that the work serve a useful purpose, and I convinced myself with a more careful reading that this one can. It is really a small linear space compendium, and can be used as a convenient reference for the beginning worker or potential worker in the field, including as it does linear space theory in Chapter 1, finite and infinite systems of integral equations in Chapter 2, linear topological spaces in Chapter 3, and infinite dimensional topology and challenge and open problems in Chapter 4. There are virtually no proofs given in the book after the first chapter, and this really brings home quickly to the reader the fact that the book is primarily a survey, or expository article. It *is* true, as the author claims, that one could use Chapter 1 as a text for a short course in strictly algebraic linear space theory, either for selected advanced undergraduate or early graduate students. If one wished a longer course, one could (as the author suggests) tack on additional material from the subsequent chapters. It could be done R. L. Moore style (with no hand-outs) or Chicago style (lots of them, and much lecturing and theorem proving by the instructor). Either way, I'm sure that this book could work out quite well as a text, *and* for the group of students the author pictures: late undergraduate and early graduate.

John von Neumann and Norbert Wiener. From Mathematics to the Technologies of Life and Death. By Steve J. Heims. M.I.T. Press, Cambridge, Mass., 1980. xviii + 547 pp., $19.95.

MARK KAC
Department of Mathematics, University of Southern California, Los Angeles, CA 90007

At first glance, this is a dual biography of two of the greatest mathematicians of our times. At the same first glance, the juxtaposition of John von Neumann and Norbert Wiener may strike one as incongruous because the two were so unlike in scientific styles and in personal characteristics.

Wiener was an analyst whose work was rooted in the traditions of the nineteenth century and who considered himself a disciple of Hardy and Littlewood. Von Neumann, on the other hand, was one hundred percent twentieth century, and, though at home in all parts of mathematics, his bent was algebraic and combinatorial.

As personalities they were even farther apart. Von Neumann was a man of the world, on easy terms with the establishment, who liked parties and good food. His view of the world, on the surface at least, was colored by a brand of easy-going cynicism traceable most likely to his being born and raised in Budapest. In his charming recollections of wartime Los Alamos ("Los Alamos from below," Cal-Tech Science and Engineering, Jan-Feb 1976) Richard Feynman credits von Neumann with implanting the idea in his head that he—Feynman—should not feel responsible for the world in which he lives. "So I have developed," says Feynman, "a very powerful sense of social irresponsibility as a result of von Neumann's advice." I shudder to think of what Heims would have done with this amusing and innocent episode had he known about it.

By contrast, Wiener was an outsider, ill at ease with and suspicious of the very establishment von Neumann found so congenial (e.g., he resigned his membership in the National Academy of Sciences). Like most prodigies, he preserved to the end of his life a number of childlike qualities which explains his total lack of cynicism; children are never cynics. It is unthinkable that Wiener would have offered a young man, even in jest, the kind of advice von Neumann offered Feynman. In his general life style, Wiener was rather ascetic (he was, e.g., a strict vegetarian); he was also unusually sensitive and easily hurt.

What Wiener and von Neumann had in common was that alone, among their contemporaries, they reached and influenced an audience immeasurably wider than that of their fellow mathematicians.

Wiener broke out through the walls of the citadel of mathematics with Cybernetics—an imprecisely defined and fuzzily circumscribed collection of facts and dreams centering around control theory as a means of understanding the behavior of machines and living beings. (My admiration for Wiener's work is too well documented to allow for any unwarranted generalizations to be drawn from my freely admitted bias against Cybernetics, the book and the doctrine.) Cybernetics was embraced almost at once by sociologists, linguists, psychoanalysts and many others. It even had an impact on religion as witness the invitation to Wiener, a few years before his death, to deliver a series of lectures at the Yale School of Divinity.* (The result was Wiener's last book, *God and Golem Inc: A comment on how Cybernetics impinges on Religion*.) To be fair, Cybernetics did have an influence on serious thinkers (e.g., the late Gregory Bateson) and it did inspire much good work in certain aspects of control theory.

Von Neumann's influence outside mathematics proper is based on contributions of a more disciplined, more technical and therefore of a less controversial nature. First, there is the Theory of Games with its profound impact on modern econometrics, and then the immense area centering around computers and automata, which von Neumann created almost singlehandedly. The crowning achievement in this area, and a veritable tour de force, was the construction (on paper)

*The Yale University Press did not publish the book; the M.I.T. Press did, and the book received one of the National Book Awards.

of self-reproducing automata which miraculously and prophetically turned out to imitate a number of biological steps involved in cell mitosis. Von Neumann also deserves a lion's share of credit for providing the impetus that led to the present-day multibillion dollar computer industry. It would thus appear that a book which would undertake a study in similarity and contrast of von Neumann and Wiener, placing them and their work in a proper historical setting, would be of great interest and value. The present book is not. In fact, it is a bad book, because the author is more interested in airing his views, opinions and prejudices than in his subjects. It is a book with a slant easily detectable from the book's subtitle *From Mathematics to the Technologies of Life and Death*, and as the story unfolds, it becomes progressively clear how, in the light of the subtitle, the author regards the "pacifist" Wiener and the "militarist" von Neumann.

Von Neumann, as is well known, was actively involved in inventing and in helping to build the atom and the hydrogen bombs. His advice on defense matters was sought at the highest levels of Government, and at his death he was one of the Commissioners of the Atomic Energy Commission. In the late forties he also briefly advocated a preventive strike against the Soviet Union. This is more than enough for Mr. Heims to make von Neumann the villain of the piece, but the editors of the M.I.T. Press should stand condemned for allowing a photograph of two deformed babies born to a survivor of the Hiroshima bombing to be included at the end of a series of innocuous family album pictures.

Freeman Dyson in reviewing the book in the February-March 1981 issue of the M.I.T. Technology Review (pp. 17-19) criticizes Mr. Heims for oversimplifying and even distorting the picture by disregarding moral issues and historical realities. But Mr. Heims does something even worse by suggesting a close connection between the political views of his protagonists and their mathematical tastes and interests.

Thus Theory of Games looms big as the purely rational substitute for a "humane" way of looking at problems ranging from economics to military strategy. More obscurely, Heims also seems to detect in von Neumann's mastery of the axiomatic method (especially as applied to foundations of Quantum Mechanics) seeds of future "hawkishness."

While admitting that "the theory of self-reproducing automata is among von Neumann's most original work" (p. 212), the author gives it a scant few lines, whereas Theory of Games and Axiomatics rate a chapter each. It is, in fact, my impression that the brief mention of the theory of self-reproducing automata was an excuse to insert a gratuitous and wholly undocumented statement that "while he had elaborate plans for developing the theory, he became sidetracked by more practical projects from computing to hydrogen weapons" (p. 212).

Mr. Heims is equally selective in his choice of Wiener's contributions for fuller treatment. Cybernetics is, of course, an obvious choice, and while it does not command a separate chapter, it permeates most of the book. By contrast Wiener's magnificent work on Tauberian theorems is given only a passing mention on p. 171 preceding a comment which will surely rate as the understatement of the century: "...and he proved new theorems pertaining to Fourier transforms in the complex plane."

Wiener's theory of Brownian motion does get a chapter of its own. But while I agree that this is Wiener's most original contribution to mathematics and science, it is, in terms of impact, quite comparable to his work in Fourier transforms in the complex domain, which is barely mentioned. In fact, most mathematicians and physicists are more familiar with the latter than with the former.

Enough has been said to conclude that Mr. Heim's book is concerned more with propaganda than with scholarship, but I am compelled to add a final footnote. On p. 427, note 30, the author states: "I am grateful for the recollections of John von Neumann's childhood recounted to me by his brothers Nicholas Vonneuman and Michael Neumann, his cousin Catherine Pedroni and his schoolmates Eugene Wigner and William Fellner." I have learned in the process of preparing this review that Mr. Nicholas Vonneuman never met Mr. Heims.

Mathematics: The Loss of Certainty. By Morris Kline. Oxford University Press, New York, 1980. 366 pp. $19.95.

RAYMOND G. AYOUB
Department of Mathematics, Pennsylvania State University, University Park, PA 16802

1. Introduction and Summary. Recent years have witnessed a resurgence of interest in the history and philosophy of mathematics, the latter attempting to elucidate the essential nature of our subject. Professor Kline has been one of the contributors and the book now under review is a continuation of his efforts. In it he amplifies some of his favorite themes but, in addition, proclaims a new dogma: since 1900, mathematics has experienced a series of crises and calamities to the point where it is now in a state of disarray with the foundations shaken to their very roots and truth an irretrievable casualty.

Professor Kline is a learned scholar and much of the book contains interesting (if not always accurate) accounts of the historical development of mathematics. It is, therefore, unfortunate that he indulges in hyperbole with such persistence and even mumpsimus.

Here in brief are his points:

(i) Present-day mathematics is a mockery of the deep-rooted and widely reputed truth and logical perfection of mathematics.
(ii) The Greek miracle and flowering of mathematics made possible the mathematization of science.
(iii) The first debacle was the discovery of non-Euclidean geometry and quaternions.
(iv) The logical basis of calculus in the 17th and 18th centuries was very tenuous and verged on the calamitous.
(v) Mathematicians of the late 19th century came close to creating a logical foundation but
(vi) the paradoxes (Russell, etc.) upset the apple cart.
(vii) Hilbert's programs held out hope for salvation but
(viii) Gödel's theorem tolled the death knell—the ultimate debacle.

Then, having nothing to do with what went before,

(ix) Applied mathematics is the only decent variety.
(x) Despite everything, mathematics is surprisingly effective.

Despite the fact that Kline does not discuss his conception of truth in mathematics, it is hard to imagine what interpretation of truth would lead to some of his conclusions.

2. Induction and Formalism. Mathematics consists in reality of two activities. The one is a mixture of intuition, induction, analogy, etc., in which certain hypothetical results are arrived at. The other is the formalism which we ascribe to the Greeks and which subjects these hypotheses to close scrutiny. The former process of discovery is not well understood, while the scrutiny of the latter has given rise to a study of mathematical logic which has led to some startling facts concerning this formalism. Both of these processes contribute, in an essential way, to the discovery and proof of mathematical results. The fact that the inductive process can be misleading should not diminish its importance nor should formalism be dismissed as mere pedantry.

For centuries, Euclidean geometry seemed to be a good model of space. The results were and still are used effectively in astronomy and in navigation. When it was subjected to the close scrutiny of formalism, it was found to have weaknesses and it is interesting to observe that, this time, it was the close scrutiny of the formalism that led to the discovery (some would say invention) of non-Euclidean geometry. (It was several years later that a satisfactory Euclidean model was devised.)

This writer fails to see why this discovery was, in the words of Kline, a "debacle." Is it not, on the contrary, a great triumph? As to quaternions, in what conceivable sense is the discovery of a noncommutative division ring a calamity? If these discoveries are disasters, what about negative

numbers and complex numbers? Had he thought of using it as an example, would Kline have (to quote a phrase Abel used in another connection) "turned away with horror" from the p-adic fields in which infinite series converge if and only if the nth term tends to zero?

3. The Decline of Formalism. To return to some of Kline's points, how can one account for the fact that in Greek mathematics, mathematical formalism was paramount, yet in the 17th and 18th centuries, calculus and its various ramifications exhibited a certain disregard for precision and rigor? This is not easy to account for, but one or two comments suggest themselves. (a) Three centuries elapsed between Thales and Euclid—it took that long to formulate a satisfactory set of axioms. Thus although Newton, for example, was perfectly aware of the deficiencies in the theory of fluxions, the ideas needed time to ripen. (b) Meanwhile the invention of calculus opened up a new and wondrous world, one which was too exciting to pause for rigor. Discoveries and applications came with great rapidity. When, as in the case of the Bernoullis and Euler, the discoveries came almost weekly, it is not too astonishing that they were not much concerned with rigor. Their discoveries had a ring of truth and that seemed sufficient unto the day. On the other hand, to condemn them and others of this epoch for their lack of rigor is simply absurd.

We come now to Kline's favorite hobby horse: non "useful" mathematics is unworthy of the pursuits of mathematicians of any but the lowest estate. It is true that he bows in the direction of Gauss, whose masterpiece, *Disquisitiones* deals with "useless" mathematics, but Kline hastily points out that Gauss contributed significantly to applied mathematics and was a professor of astronomy. He does not explain why the determination of the orbit of Ceres is more important than the quadratic law of reciprocity.

Now Kline has raised tacitly an important question. In the vast outpouring of mathematics, how is one to distinguish the good from the bad, the significant from the insignificant? In the search for an easy answer Kline and others have hit upon a simple principle: if it is "useful," it is good. To this principle, I. Shafarevich has a response. He quotes from the gospel according to Thomas: "...if the spirit come for the sake of the flesh, it is a miracle of miracles." "All of the history of mathematics," writes Shafarevich, "is a convincing proof that such a 'miracle of miracles' is impossible."

4. The Elusive Nature of Mathematics. To make aesthetic or scientific judgments about mathematics, we should seek to understand its nature. Not much has been written on this question and what has been written is unsatisfying. Various authors, Aristotle, Benjamin Peirce, A. N. Whitehead, Roger Bacon, have given definitions of mathematics, but so unsatisfying are these that the best is that of Bacon who writes that "mathematics is the study which makes men subtile." This unsatisfactory situation is to be contrasted with other disciplines. If you ask a biologist about the nature and objectives of biology, his answer would be simple: to understand the nature of organisms and of life and to apply this understanding. All biology is directed to these objectives. A mathematician can give no such clear-cut answer.

We see, therefore, a fundamental difference between our subject and other sciences. In other sciences the essential problems are forced upon the subject from external sources, and the scientist has no control over the ultimate end. The mathematician, however, is free to prescribe not only the means of realizing the end, but also the end itself. (Can this be why mathematicians are so rebellious and find external restraint abhorrent?) Despite this apparent anarchy, our subject thrives and prospers. It is like a beautiful garden with a profusion of shrubs, flowers and trees and a host of gardeners. And if a gardener wishes to cultivate a new strain, he is completely at liberty to do so. (Of course he may be constrained by funding agencies!) Given the proper nourishment, his plot could flower into one of the most prominent features of the garden. Consider as an example, the problem of counting solutions of congruences. In the hands of a galaxy of stars, beginning with Gauss and including Jacobi, Artin, Hasse, Weil, Dwork, Grothendieck and Deligne, this has flowered into one of the glories of our garden.

There are, unfortunately, difficulties to which this individualism leads. Leaving the simile, the questions are

(i) How does one choose the problems?
(ii) Having chosen problems, how can we predict which will have significance?
(iii) How or who is to judge this significance, and how or who is to pass upon the beauty or importance of a segment of mathematics?

Kline has, as noted above, an easy answer—applied mathematics. But let us look as a couple of examples:

(a) The Greeks made famous the problem of perfect numbers and while it can be claimed that they provided the impetus for Euclid's proof of unique factorization, two millenia of work has failed to produce further significant mathematics arising from this problem.
(b) Fermat's last theorem is but one of a plethora of diophantine equations, yet it led to cyclotomy, algebraic number theory, ideal theory, etc. To be sure cyclotomy also came from other sources, higher reciprocity, division points, etc.

One problem (a) led to a dead end while the other (b) is still providing stimulus for beautiful new directions. Is it because (b) attracted the attention of more creative mathematicians, especially Kummer, etc.? Hardly, since both Euclid and Euler contributed to (a). Could one have predicted the significance of the one and not the other? Alas, not even the great Gauss foresaw the importance of (b). Is (a) then doomed to remain in the backwaters of mathematics? It is impossible to say.

Despite these uncertainties, and despite some differences of opinion, it is a startling fact that the amount of agreement on what constitutes great mathematics far outweighs the amount of disagreement. This fact is one of the mysteries to which no adequate answer has yet been given.

Kline is highly critical of the outpouring of what some might call "contrived mathematics." While he has a point, his criticism is unfair for his perspective is distorted—we learn from the past only what has been deemed of high calibre. After all, who now remembers or cares about Molweide's formulae? To be sure, it is a depressing fact to many that much of what is now produced will be cast into oblivion, but does that mean that mathematicians should abandon their interests and turn to mathematics of the "'flesh"? Kline cannot seriously think so.

5. Mathematics and the Real World. Does mathematics describe the "real" world? The answer is that mathematics has been astonishingly successful in modelling some aspects of the real world. Indeed, E. Wigner has commented on the unreasonable effectiveness of mathematics in describing real world phenomena. Kline acknowledges its effectiveness and utility.

On the other hand, just because Euclidean geometry can no longer be regarded as the correct abstraction of the geometry of space is certainly no grounds for bemoaning the loss of "truth." Truth about the real world is arrived at by stages. There is nothing wrong with the model of planetary motion given by epicycles so long as observations fit the predicted motion. There is nothing wrong with using Euclidean geometry and its derivative spherical trigonometry in navigation. It serves the purpose adequately. The fact that Euclidean geometry does not describe the global universe is no basis for Kline to proclaim disaster.

Finally, Professor Kline does not deal honestly with his readers. He is a learned man and knows perfectly well that many mathematical ideas created in abstracto have found significant application in the real world. He chooses to ignore this fact, acknowledged by even the most fanatic opponents of mathematics. He does this to support an untenable dogma. One is reminded of the story of the court jester to Louis XIV: the latter had written a poem and asked the jester his opinion. "Your majesty is capable of anything. Your majesty has set out to write doggerel and your majesty has succeeded." On balance, such, alas, must be said of this book.

INDEX TO VOLUME 89, 1982

THE AMERICAN MATHEMATICAL MONTHLY

AUTHOR INDEX

INDEX OF SPECIAL RECURRING TOPICS

SUBJECT INDEX

This index uses the current version of AMS 1980 Mathematics Subject Classification. Underlined entries indicate that the subject heading is secondary.

MISCELLANEA

PROBLEMS AND SOLUTIONS

PROBLEMS PROPOSED

PROBLEMS SOLVED

SOLUTIONS

Numbers in boldface type refer to problems; those in lightface, to pages.

REVIEWS

Names of authors are in ordinary type; those of reviewers, in capitals.

FILMS

ERRATA

Vol. 89

Kenschaft, "Black Women in Mathematics in the United States," October, 1981: In attempting to list all American Black women with doctorates in mathematics, the author missed two women. *Annie Marie Garraway* received her Ph.D. from the University of California at Berkeley in 1967 and is now employed by Bell Laboratories in Columbus, Ohio. *Emma Rose Fenceroy* received her degree from the University of Alabama in 1979 and is now head of the Department of Mathematics at Florida A & M University in Tallahassee. Three more Black women have completed their degrees since the article went to press.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 89, Number 10 December 1982

Contents

(ISSN 0002-9890)

NOTICE TO AUTHORS

See statement of editorial policy (this volume, p. 3).

Follow the format in current issues. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Supply the Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. Submit material to the editor or appropriate associate editor as indicated below.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $20 included as part of the annual dues of $40. Students receive a 50% discount. The library subscription price is $50 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

SOLUTION OF TWO DIFFICULT COMBINATORIAL PROBLEMS WITH LINEAR ALGEBRA

ROBERT A. PROCTOR*
Department of Mathematics, University of California, Los Angeles, CA 90024

(i) *Pick six positive real numbers—any six positive real numbers. If you chose*

$$1, e, \pi, 4, \sqrt{67}, 98.6,$$

you're in trouble, because the object of this game is to have as many subsets as possible adding up to the same sum. For example, choosing

$$5, 6, 8, 9, 13, 14$$

yields three subsets with the same sum:

$$14 = 5 + 9 = 6 + 8.$$

The set

$$3, 4, 5, 6, 7, 8$$

has four subsets adding up to the same number:

$$7 + 8 = 3 + 4 + 8 = 3 + 5 + 7 = 4 + 5 + 6.$$

Choosing

$$1, 2, 3, 4, 5, 6$$

yields five subsets with the same sum:

$$4 + 6 = 1 + 3 + 6 = 1 + 4 + 5 = 2 + 3 + 5 = 1 + 2 + 3 + 4.$$

Is this the best possible? This problem can be more generally posed for n positive real numbers: Then it seems that no collection does better than

$$1, 2, 3, \ldots, n - 1, n.$$

Call this problem the **subset sum problem**.

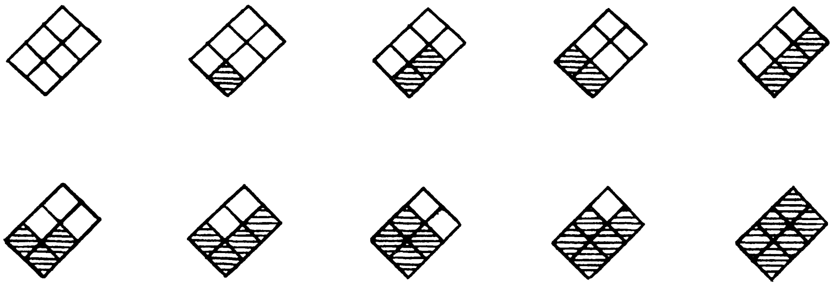


FIG. 1.

(*ii*) *Fix two positive integers m and n. Draw an $m \times n$ grid of squares "on tilt" as in Fig.* 1. *Now shade in some of the squares so that there are no unshaded squares below shaded squares—i.e., so that if the shaded squares were blocks in a rectangular frame, none would slide down. Call such a*

Author's autobiography: My research centers on the combinatorial study of discrete structures arising in Lie theory. When mathematics permits, my hobbies include country swing dancing, soccer, and surfing. I received my graduate education primarily at M.I.T., where my thesis adviser was Richard P. Stanley. I was also a visiting graduate student at the University of California at San Diego for one year, and this is when the ideas contained in this article arose. I am now a Hedrick Assistant Professor at the University of California at Los Angeles.

*Supported in part by an NSF Graduate Fellowship.

shading a **proper shading**. *As an index k runs from 0 to mn, count how many proper shadings there are with k shaded squares. For example, if $m = 2$ and $n = 3$, then the count is*

$$1,1,2,2,2,1,1$$

as k runs from 0 to 6. For $m = 2$ and $n = 4$, the count is

$$1,1,2,2,3,2,2,1,1.$$

And for $m = 3$ and $n = 4$ one counts

$$1,1,2,3,4,4,5,4,4,3,2,1,1.$$

It seems that the count always weakly increases until half of the squares have been shaded, and then weakly decreases until all of the squares have been shaded. In other words, there seem to be no "dips" in the count. It gets bigger, then smaller. Is this always true for any size grid? Call this the **grid shading problem.**

1. Introduction. *Letting the cats out of the bags, the answers to both of the questions above are what you would expect. The choice*

$$1,2,3,\ldots,n-1,n$$

is the best possible, and there are not any dips in the proper shading count for any values of m and n. Proving either of these answers correct is surprisingly hard. All known proofs of these results involve representations of Lie algebras or the symmetric group in some form. We'll give proofs phrased entirely in terms of elementary linear algebra. These proofs were obtained by translating the essential parts of Lie algebraic proofs into linear algebra. So knowledge of undergraduate linear algebra is the only background you'll need.

Let $p_0, p_1, \ldots, p_r$ be a sequence of numbers. If

$$p_0 \leqslant p_1 \leqslant \cdots \leqslant p_{h-1} \leqslant p_h \geqslant p_{h+1} \geqslant \cdots \geqslant p_{r-1} \geqslant p_r$$

for some h between 0 and r, then the sequence is said to be **unimodal**. The grid shading problem can now be succinctly stated: *Is the sequence of proper shading counts unimodal for all values of m and n?*

How did this problem and its solution arise? Unfortunately from a theatrical point of view, the grid shading problem first arose in the subject in which it was ultimately solved: Classical Invariant Theory. (Not a dashing rescue of some befuddled combinatorialists by the Lie algebra cavalry: The ancestors of modern day Lie representation theorists got themselves into and out of this one!) While finding all covariants of a binary quantic in the early 1850's, Arthur Cayley apparently took the unimodality of the grid shading counts completely for granted. At the same time, he accepted without proof the independence of a certain set of linear equations. In the words of James Sylvester, this crucial gap in Cayley's methods remained for "upwards of a quarter century." Then in 1878, "by aid of a construction drawn from the resources of Imaginative Reason," Sylvester "accomplished with scarcely an effort a task which (he) had believed lay outside the range of human power," and showed that Cayley's equations were in fact independent. The unimodality of the proper shading counts is an immediate consequence of the independence of these equations. Other proofs have appeared over the last century in various contexts, including representations of the symmetric group, Hodge theory, and representations of Lie algebras. But all of these proofs, including the one presented in this article, are related to the methods of Cayley and Sylvester. However, this being 1982, instead of using "a construction drawn from the resources of Imaginative Reason," we'll use a "trick" for our crucial step.

Upon hearing of this unimodality theorem and its mysterious proofs a few years ago, some combinatorialists decided to look for a more satisfying "combinatorial" proof: For each $k < mn/2$ (less than half of the squares shaded), explicitly describe a one-to-one map from the set of proper shadings with k squares shaded into the set of proper shadings with $k + 1$ squares shaded. This would imply that the proper shading count increases until half of the squares have been shaded. A

simple symmetry argument then completes the proof. At first glance it seems like it should be easy to find such a proof, but to date such proofs have been found only for the cases where one of m or n is less than or equal to 5.

In contrast to the grid shading problem, the story behind the subset sum problem *does* have something of a dramatic twist. Since this problem was of genuinely combinatorial origins, its solution was almost necessarily by accident: No one in his right mind would consider using the cohomology of projective algebraic varieties! However, this was the technique which led to the first solution of this problem. First proposed by that grandmaster of intriguing combinatorial problems himself, Paul Erdős, and Leo Moser in 1963, no progress was made on the problem until 1969. At that time, Bernt Lindström reduced the problem to showing that a certain family of partially ordered sets have the "Sperner" property. Completely unaware of both the original problem and Lindström's reduction, Richard Stanley showed in 1978 that this family of partially ordered sets do have the Sperner property. Stanley's methods used the hard Lefschetz theorem of algebraic geometry, which concerns the effect of multiplication by the cohomology class corresponding to a hyperplane section in the cohomology ring of a projective algebraic variety. However, the solution was not complete until these two pieces were combined. The connection was made by Larry Harper during a phone call with Stanley, the original purpose of which was to discuss a house sublet during a sabbatical leave! The Lie algebraic solution from which our proof is derived is closely related to Stanley's algebraic geometric solution.

The solutions presented here proceed as follows: We'll associate a family of partially ordered sets to each problem, and then show how the questions at hand can be translated into properties of the partially ordered sets. Next we'll associate vector spaces with linear operators to each partially ordered set. Everything will then be reduced to showing that the linear operators have the largest ranks possible. The solutions will be completed by proving this with a trick borrowed from the representation theory of Lie algebras.

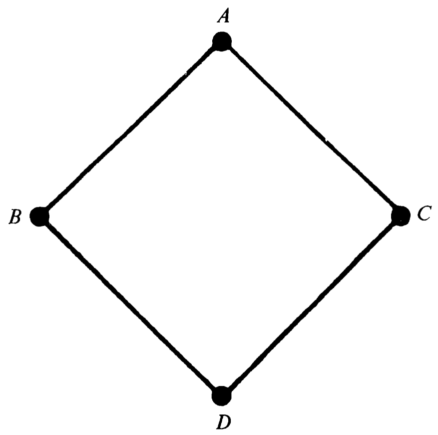


FIG. 2.

2. Poset Formulation. A **finite partially ordered set** ("**poset**" for short) is just a finite set upon which an ordering relation $\leqslant$ has been defined. As indicated by the word "partial," it is not necessary for any two elements to be related by $\leqslant$. For example, Fig. 2 shows what a poset describing the logical dependence of four chapters in a textbook might look like: Chapter A must precede Chapters B and C, and Chapters B and C must precede Chapter D, but neither Chapter B nor Chapter C must precede the other.

For any fixed values of m and n, there's a natural partial ordering of the various proper shadings of an $m \times n$ grid: Order the proper shadings "by containment." See Fig. 3 for the case $m = 2$ and $n = 2$. Given a proper shading, let a_1 be the number of squares shaded in its top right

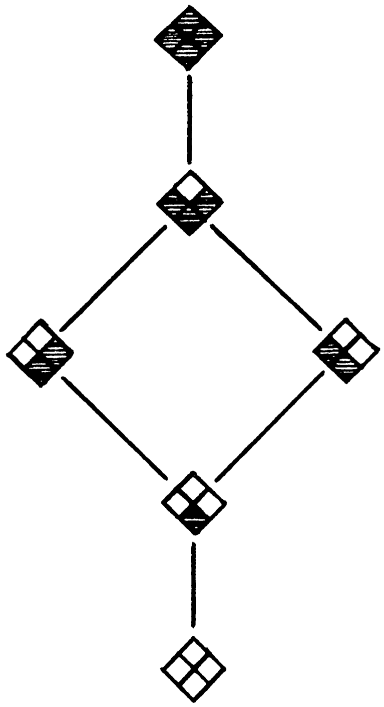


FIG. 3.

row, a_2 the number shaded in the next row,..., and a_n the number shaded in the bottom left row. Note that

$$0 \leqslant a_1 \leqslant a_2 \leqslant \cdots \leqslant a_{n-1} \leqslant a_n \leqslant m.$$

Let the n-tuple

$$\mathbf{a} = (a_1, a_2, \ldots, a_n)$$

denote this shading. It's easy to see that there's a 1-1 correspondence between the collection of all such n-tuples and the set of all proper shadings of the $m \times n$ grid. Now the poset of proper shadings can be officially described:

$$\mathbf{a} \leqslant \mathbf{b}$$

if and only if

$$a_1 \leqslant b_1, \quad a_2 \leqslant b_2, \quad \ldots, \quad a_n \leqslant b_n.$$

We'll call this poset $\boldsymbol{L(m, n)}$. Fig. 4 shows $L(3,3)$.

Whenever one element of a poset lies immediately above another element, we'll say that the first element **covers** the second. For example, the element $(1,2,3)$ covers the element $(1,1,3)$ in $L(3,3)$. A poset P which can be split up into $r+1$ subsets

$$P_0, P_1, P_2, \ldots, P_{r-1}, P_r$$

such that elements in P_k cover *only* elements in P_{k-1} is called **ranked**. (The smallest example of a poset which cannot be ranked has a diagram which is a pentagon.) If we let p_k denote the number of elements in the $\boldsymbol{k}$**th rank** P_k, the sequence of numbers

$$p_0, p_1, p_2, \ldots, p_{r-1}, p_r$$

lists the sizes of the ranks of P. If this sequence is unimodal, we'll call P a **rank unimodal** poset.

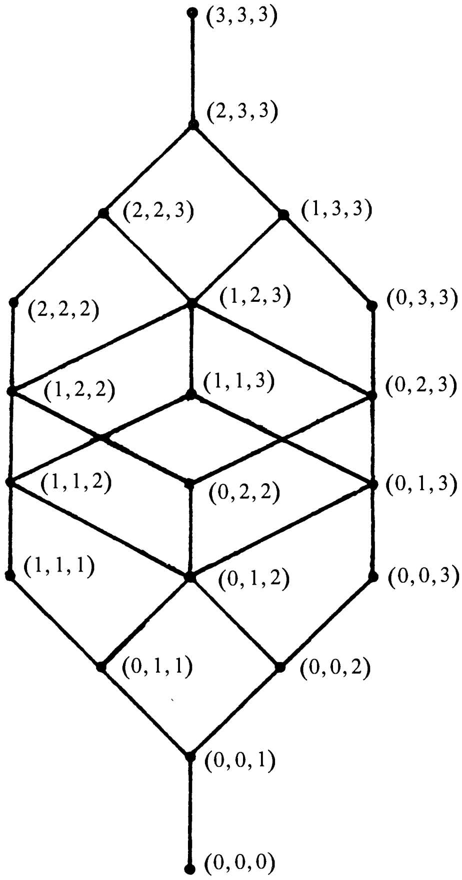


FIG. 4. $L(3,3)$.

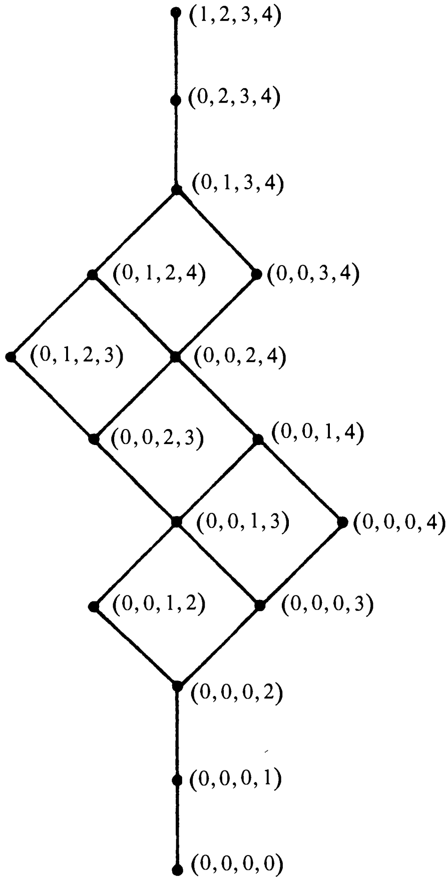


FIG. 5. $M(4)$.

In the posets $L(m, n)$, the 0th rank consists of just the empty grid. As we generate all proper shadings by adding one shaded square at a time, note that all the grids with k squares shaded will lie in the kth rank of $L(m, n)$. So the size of the kth rank of $L(m, n)$ is the number of proper shadings of an $m \times n$ grid with k squares shaded. Therefore the grid shading problem can be restated as: *Show that the posets $L(m, n)$ are rank-unimodal for all values of m and n.*

There's a family of posets closely related to the $L(m, n)$ whose structures play a crucial role in the solution of the subset sum problem. Let ***M(n)*** denote the set of all n-tuples of integers

$$\mathbf{b} = (b_1, b_2, \ldots, b_n)$$

such that

$$0 = b_1 = \cdots = b_j < b_{j+1} < b_{j+2} < \cdots < b_n \leqslant n$$

where j is some integer $0 \leqslant j \leqslant n$. For example, if $n = 3$, there are eight such 3-tuples:

$$(0,0,0), \quad (0,0,1), \quad (0,0,2), \quad (0,0,3),$$
$$(0,1,2), \quad (0,1,3), \quad (0,2,3), \quad (1,2,3).$$

Now define a partial ordering on the set $M(n)$ exactly as we did for $L(m, n)$:

$$\mathbf{a} \leqslant \mathbf{b}$$

if and only if

$$a_1 \leqslant b_1, \quad a_2 \leqslant b_2, \quad \ldots, \quad a_n \leqslant b_n.$$

See Fig. 5 for a picture of $M(4)$.

The 0th rank of $M(4)$ consists of the 4-tuple $(0,0,0,0)$. As we move upward in the poset, we see that each succeeding rank consists of 4-tuples whose entries add up to a sum 1 greater than the sums of the 4-tuples in the preceding rank. Therefore the kth rank of $M(4)$ consists of 4-tuples which add up to k. For example, the 5th rank of $M(4)$ consists of $(0,0,2,3)$ and $(0,0,1,4)$. It is not hard to see that this is true for the ranks of any $M(n)$.

Now recall that the subset sum problem concerned the sums of the elements of subsets of a set of n positive real numbers. In particular, it was claimed that the set

$$N = \{1,2,3,\dots, n-1, n\}$$

had at least as many subsets adding up to a common sum as did any other set of n positive real numbers. If we look back at our definition of $M(n)$, we see that there is a 1-1 correspondence between its n-tuples and the subsets of the set N. Note that under this correspondence, the entries of the n-tuple add up to the same sum as the elements of the subset do. So two subsets of N will add up to the same number k if and only if both of their corresponding n-tuples lie in the kth rank of $M(n)$. Therefore the largest collection of subsets of N with equal sums will be found by taking the subsets which correspond to the n-tuples lying in the largest rank of $M(n)$.

Now remember that what we want to show is that no other set S of n positive real numbers will do better than the set N. Take any such set S and list its elements in increasing order:

$$S = \{s_1, s_2, s_3, \dots, s_{n-1}, s_n\},$$
$$s_1 < s_2 < s_3 < \cdots < s_{n-1} < s_n.$$

If

$$A = \{s_{i_1}, s_{i_2}, \dots, s_{i_q}\}, \quad i_1 < i_2 < \cdots < i_q,$$

is a subset of S, then define

$$\mathbf{a} = (0,0,\dots,0,i_1,i_2,\dots,i_q)$$

to be a corresponding element of $M(n)$. It will no longer be true for S (as it was for N) that two subsets corresponding to n-tuples in the same rank of $M(n)$ will have equal sums. However, the following *will* be true:

$$\mathbf{a} < \mathbf{b} \quad \text{in } M(n)$$

implies that

$$\Sigma(A) < \Sigma(B),$$

where $\Sigma(A)$ and $\Sigma(B)$ denote the sums of the subsets of S corresponding to the n-tuples $\mathbf{a}$ and $\mathbf{b}$. To see this, just write the elements of S above the entries of $\mathbf{a}$ and $\mathbf{b}$ to which they correspond. For example, if

$$S = \{\sqrt{2}, 11, 10\pi, 55\}$$

and

$$A = \{1,3\} \quad \text{and} \quad B = \{1,2,4\},$$

then write

$$\begin{matrix} & & \sqrt{2} & 10\pi \\ (0, & 0, & 1, & 3) \end{matrix} \quad \text{and} \quad \begin{matrix} & \sqrt{2} & 11 & 55 \\ (0, & 1, & 2, & 4). \end{matrix}$$

Observe that the element of S written above the ith entry of $\mathbf{a}$ will be less than the element of S written above the ith element of $\mathbf{b}$. Therefore the sum of the elements in A will be less than the sum of the elements in B.

The upshot of all this is: *In order for two subsets of S to have equal sums, they must correspond to incomparable elements of $M(n)$*. (Two elements x and y of a poset are said to be **incomparable** if $x \nleqslant y$ and $y \nleqslant x$.) This tells us that the largest collection of subsets of S with equal sums can be no larger than the largest collection of mutually incomparable elements of $M(n)$. A little experimentation for small values of n seems to indicate that the largest collection of mutually incomparable elements of $M(n)$ is never larger than the biggest rank of $M(n)$. If we could prove this to be true for all n, then we'd be done, as the following summary indicates:

Largest collection of subsets of S with equal sums
is no larger than the
Largest collection of mutually incomparable elements of $M(n)$
which is conjectured to be no larger than the
Largest rank of $M(n)$
which is equal in size to the
Largest collection of subsets of N with equal sums.

Any ranked poset which has no collection of mutually incomparable elements larger than its largest rank is said to be **Sperner**. The above argument has reduced the subset sum problem to: *Show that the posets $M(n)$ are Sperner for all values of n.*

The methods we'll use in this article will actually show all $L(m, n)$ and $M(n)$ to be both rank unimodal and Sperner. Here's one way to show that a ranked poset P has these properties. Suppose the ranks of P are

$$P_0, P_1, P_2, \dots, P_{r-1}, P_r$$

and suppose we want to show that the sequence of rank sizes is unimodal with a peak at $k = h$. Assume for the time being that to each element x in a rank P_k with $k < h$, we can assign an element y from the rank P_{k+1} such that $x < y$ and such that no element y from the rank P_{k+1} is used more than once when all of the elements of rank P_k have been taken care of. If we can do this, we'll say that we have a **matching** of rank P_k into rank P_{k+1}. Also suppose that we can do the reverse process for each pair of ranks P_k and P_{k-1} when $k > h$ (i.e., find a 1-1 map from P_k into P_{k-1} such that if w in P_{k-1} is assigned to x in P_k, then $w < x$). The following lemma asserts that we will have solved both of our problems if we can find such matchings for $L(m, n)$ and $M(n)$.

LEMMA. *Let P be a ranked poset with $r + 1$ ranks. If rank P_k can be matched into rank P_{k+1} for $k < h$ and rank P_k can be matched into rank P_{k-1} for $k > h$, then the poset P is rank unimodal and Sperner.*

Proof. As before, let p_k be the number of elements in the kth rank. The matching conditions clearly imply that $p_k < p_{k+1}$ for $k < h$ and that $p_{k-1} > p_k$ for $k > h$. Thus the sequence of rank sizes is unimodal about $k = h$.

Picture the matchings as special edges in the diagram of P. (See Fig. 6.) Glue the various matchings together and obtain "chains" of elements of the poset which stream into the rank P_h from above, and below. (A **chain** is a subset of a poset in which any two elements are related by $\leqslant$. Chains can be visualized as ascending or descending sequences of elements in the diagram of the poset.) Every element of P lies on one of these chains. Now take any set of mutually incomparable elements in P and plot them on the diagram of P. Each of these elements lies on one of the chains. By the definition of incomparable, no two can lie on one chain. But there are exactly as many chains as there are elements in the biggest rank P_h. Therefore there can't be more elements in this mutually incomparable subset than there are elements in the largest rank P_h. So the poset P must be Sperner, as desired.

3. Linear Algebra Formulation. So far we have translated the original two problems into the question of whether certain matchings exist in the posets $L(m, n)$ and $M(n)$. The proof of the

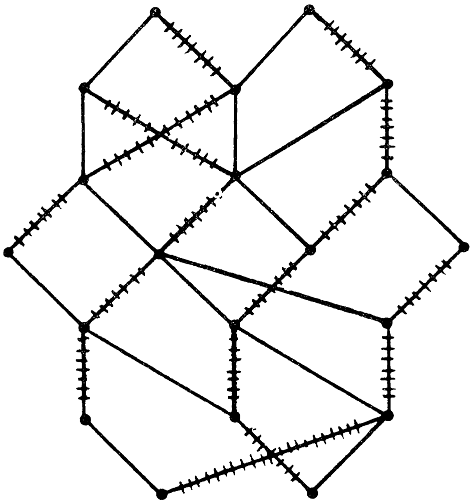


FIG. 6.

existence of such matchings, the heart of the solutions, is in the next section. To get there from here, we must first describe the role that linear algebra plays. In this section we'll work with an arbitrary ranked poset P with ranks

$$P_0, P_1, P_2, \dots, P_{r-1}, P_r.$$

Suppose that the elements of P are $a, b, \dots, e$. Let $\tilde{P}$ denote the vector space over the complex numbers which has a basis consisting of vectors $\tilde{a}, \tilde{b}, \dots, \tilde{e}$ which correspond to the elements of P. Under this set-up, $\tilde{P}_k$ will denote the subspace of $\tilde{P}$ spanned by those basis elements which correspond to elements in rank P_k. We'll call $\tilde{P}_k$ the **kth rank subspace** of $\tilde{P}$. Note that

$$\dim \tilde{P}_k = p_k,$$

the number of poset elements in the rank P_k.

Now define the **order operator** of P to be the linear operator X on $\tilde{P}$ given by:

$$X\tilde{a} = \Sigma_{b \text{ covers } a} \tilde{b}.$$

Unofficially: When X acts on a poset element, it produces the sum of poset elements covering that element. Note that

$$X(\tilde{P}_k) \subseteq \tilde{P}_{k+1}.$$

So let's define X_k to be the linear transformation from $\tilde{P}_k$ to $\tilde{P}_{k+1}$ obtained by restricting X to $\tilde{P}_k$. The matrix for X_k with respect to the poset element bases for $\tilde{P}_k$ and $\tilde{P}_{k+1}$ is a $p_{k+1} \times p_k$ matrix consisting of 0's and 1's. The locations of the 1's describe which elements of P_{k+1} cover which elements of P_k.

We're now ready to state the next step of the solutions:

LEMMA. *Let P be a ranked poset with $r+1$ ranks and order operator X. If there is some h such that the X_k are injective for $k < h$ and surjective for $k \geqslant h$, then P is rank unimodal and Sperner.*

Proof. Suppose $k < h$. Then X_k is injective and its matrix with respect to the poset basis has rank p_k. This matrix must have at least one nonzero $p_k \times p_k$ minor. The usual determinantal expansion of this minor into $p_k!$ terms must have at least one nonzero term. This term is the product of p_k 1's, where no two of these 1's lie in the same row or the same column of the matrix for X_k. Now the columns of the matrix for X_k are indexed by the elements of P_k and the rows by elements of P_{k+1}. Using the nonzero term of the nonzero minor, we can assign to each element in P_k a covering element from P_{k+1} in such a way that no element from P_{k+1} is used twice. In other

words, we get a matching of P_k into P_{k+1}. The same method produces matchings of P_k into P_{k-1} when $k > h$ and X_{k-1} is surjective. Apply the lemma of the previous section to finish the proof.

4. Order Operators Have Maximal Ranks. In this section we'll complete the solutions of the two problems by showing that the order operators for the posets $L(m, n)$ and $M(n)$ satisfy the requirements of the lemma above. The construction used will seem somewhat unmotivated and *ad hoc* to anyone unfamiliar with representations of the Lie algebra $sl(2, \mathbb{C})$. However, the existence of such a construction is almost necessary in these circumstances—more on this in the last section.

Recall that the elements of $L(m, n)$ and $M(n)$ are being denoted by n-tuples

$$\mathbf{a} = (a_1, a_2, \ldots, a_n).$$

Also recall that the order operator X for any poset is defined by

$$X\tilde{\mathbf{a}} = \Sigma_{\mathbf{b} \text{ covers } \mathbf{a}} \tilde{\mathbf{b}}.$$

We'll need two other linear operators H and Y on the vector spaces $\tilde{L}(m, n)$ and $\tilde{M}(n)$. Set

$$H\tilde{\mathbf{a}} = [2(a_1 + a_2 + \cdots + a_n) - mn]\tilde{\mathbf{a}}$$

on $\tilde{L}(m, n)$, and

$$H\tilde{\mathbf{a}} = \left[2(a_1 + a_2 + \cdots + a_n) - \frac{n(n+1)}{2}\right]\tilde{\mathbf{a}}$$

on $\tilde{M}(n)$. Oh yes—we forgot to mention in Section 2 that if $\mathbf{b}$ covers $\mathbf{a}$ in either $L(m, n)$ or $M(n)$, then there is some index i such that $a_i = b_i - 1$ and $a_j = b_j$ when $j \neq i$. But no harm done, since only now do we need this fact: Define the third operator Y on either $L(m, n)$ or $M(n)$ by

$$Y\tilde{\mathbf{b}} = \Sigma_{\mathbf{b} \text{ covers } \mathbf{a}} c(\mathbf{a}, \mathbf{b})\tilde{\mathbf{a}},$$

where (assuming $a_i = b_i - 1$)

$$c(\mathbf{a}, \mathbf{b}) = (m + n - a_i - i)(a_i + i)$$

for $L(m, n)$, and $c(\mathbf{a}, \mathbf{b}) = n(n+1)/2$ if $a_i = 0$, otherwise $= (n - a_i)(n + a_i + 1)$ for $M(n)$. In passing, we culturally remark that each set of three operators defines a representation of $sl(2, \mathbb{C})$ on $\tilde{L}(m, n)$ or $\tilde{M}(n)$.

If you picture the posets $L(m, n)$ and $M(n)$ as in Figures 4 and 5, then the action of X raises vectors by one level, the action of H leaves vectors in their original levels, and the action of Y lowers vectors by one level. By now it should be apparent that the two cases $L(m, n)$ and $M(n)$ are very similar. To save ink, we'll usually refer only to $L(m, n)$ from now on. The rank subspaces of $\tilde{L}(m, n)$ will be denoted by

$$\tilde{L}_0, \tilde{L}_1, \ldots, \tilde{L}_k, \ldots, \tilde{L}_{mn}.$$

Now that we've dispensed with the reductions of the problems, definitions, notation, and preliminary observations, we're ready to get to work. What do we want to prove? Remember that X_k denotes the restriction of X to the rank subspace $\tilde{L}_k$ (or $\tilde{M}_k$ for $\tilde{M}(n)$).

LEMMA. *The linear transformations X_k for $\tilde{L}(m, n)$ are injective if $k < mn/2$ and surjective if $k \geqslant mn/2$. The linear transformations X_k for $\tilde{M}(n)$ are injective if $k < n(n+1)/4$ and surjective if $k \geqslant n(n+1)/4$.*

Once we prove this, the solutions are complete. The proof consists of two parts. The first part concerns the "commutation relations" between the three operators X, Y, and H. The second part constructs new bases for the vector spaces $\tilde{L}(m, n)$ and $\tilde{M}(n)$.

Getting on with the first part, we first claim that

$$HX - XH = 2X \quad \text{and} \quad HY - YH = -2Y.$$

These aren't hard to prove: Take $\mathbf{a}$ in L_k. Then

$$H\tilde{\mathbf{a}} = (2k - mn)\tilde{\mathbf{a}} \quad \text{and} \quad H(X\tilde{\mathbf{a}}) = (2k + 2 - mn)X\tilde{\mathbf{a}},$$

since $X\tilde{\mathbf{a}}$ is in $\tilde{L}_{k+1}$. Therefore

$$[HX - XH]\tilde{\mathbf{a}} = [2k + 2 - mn - (2k - mn)]X\tilde{\mathbf{a}} = 2X\tilde{\mathbf{a}}.$$

And similarly for the relation $HY - YH = -2Y$.

More work is required to prove the third commutation relation:

$$XY - YX = H.$$

Again look at the effect of the operators on a basis vector $\tilde{\mathbf{a}}$ in $\tilde{L}_k$. Both $XY\tilde{\mathbf{a}}$ and $YX\tilde{\mathbf{a}}$ lie in $\tilde{L}_k$ and are thus linear combinations of the vectors $\tilde{\mathbf{d}}$ with $d_1 + d_2 + \cdots + d_n = k$. In terms of n-tuples, the effect of X is roughly described as adding 1 to each component of the n-tuple at a time. And Y subtracts 1 from each component at a time. So $\tilde{\mathbf{d}}$ appears in the expansion for $[XY - YX]\tilde{\mathbf{a}}$ exactly whenever there are indices i and j such that

$$d_i = a_i - 1 \quad \text{and} \quad d_j = a_j + 1.$$

If $i \neq j$, then the resulting terms in both $XY\tilde{\mathbf{a}}$ and $YX\tilde{\mathbf{a}}$ have coefficient $(m + n - a_i - i + 1)(a_i + i - 1)$ and thus cancel each other. Therefore $[XY - YX]\tilde{\mathbf{a}}$ is a scalar multiple of $\tilde{\mathbf{a}}$: the only contribution comes from adding then subtracting, or subtracting then adding 1 to the same component. Explicitly,

$$[XY - YX]\tilde{\mathbf{a}} = \Bigg\{ \sum_{\substack{1 \leqslant i \leqslant n \\ a_{i-1} < a_i}} (m + n - a_i - i + 1)(a_i + i - 1) - \sum_{\substack{1 \leqslant i \leqslant n \\ a_i < a_{i+1}}} (m + n - a_i - i)(a_i + i) \Bigg\} \tilde{\mathbf{a}}.$$

The summation conditions result from the requirement that the n-tuples must always satisfy

$$0 \leqslant a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n \leqslant m.$$

(And we have set $a_0 = 0$ and $a_{n+1} = m$.) A close look at these summations reveals that it's O.K. to drop the extra conditions on the summations—any new terms appearing will be zero or will cancel each other. So then

$$\begin{aligned}[XY - YX]\tilde{\mathbf{a}} &= \Bigg\{ \sum_{1 \leqslant i \leqslant n} [(m + n - a_i - i + 1)(a_i + i - 1) - (m + n - a_i - i)(a_i + i)] \Bigg\} \tilde{\mathbf{a}} \\ &= [2(a_1 + a_2 + \cdots + a_n) - mn]\tilde{\mathbf{a}} \\ &= H\tilde{\mathbf{a}}.\end{aligned}$$

Similar steps work for $M(n)$.

This brings us to the second part of the proof of the lemma. We are now ready to take advantage of one of the nicer features of linear algebra, the ability to change bases. Although changing bases will scramble the information currently contained in the matrix descriptions of the transformations X_i, the new basis will help us show that these transformations have the correct ranks. The lemma in Section 3 then provides for the unscrambling of the information back to its original form with the knowledge that the necessary matchings exist.

The construction given below replaces the poset basis used until now for $\tilde{L}(m, n)$ with a new basis which is depicted schematically in Fig. 7. Interpret this diagram as follows. The dots in the kth level of the diagram represent new basis vectors lying in the kth rank subspace $\tilde{L}_k$. And the new basis vectors lying in a vertical line are related by the recursion

$$w_{i+1} = Xw_i.$$

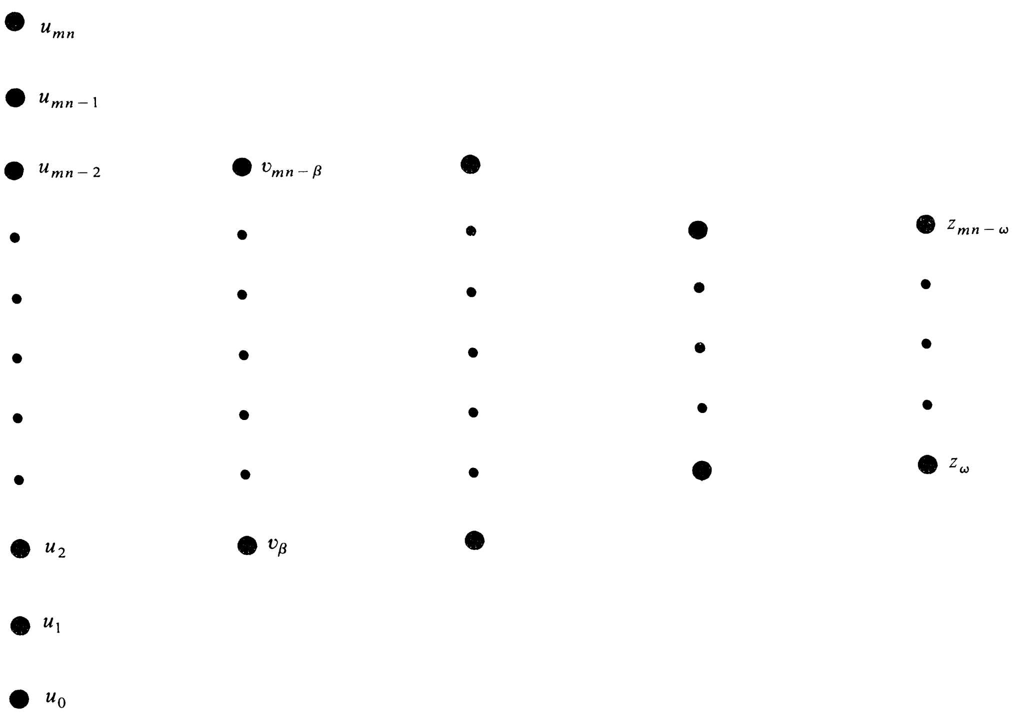


FIG. 7. New basis for $\tilde{L}(m, n)$.

Any sequence of vectors related in this manner is called a **string** of vectors. Our new basis will consist of a collection of strings of vectors with each string symmetric about the middle subspace(s).

Let the first member u_0 of the new basis be $\tilde{\mathbf{a}}_0$, where $\mathbf{a}_0$ is the only element of L_0. Let U be the subspace of $\tilde{L}(m, n)$ spanned by the various vectors obtained by acting on u_0 with all possible compositions of the operators X, Y, and H. For example, $5XYu_0 - 7XHXXu_0$ is a typical element of U. By repeatedly using the commutation relations

$$HX = XH + 2X, YH = HY + 2Y \quad \text{and} \quad YX = XY - H,$$

it's possible to express any composition of the operators X, Y, and H as a linear combination of terms of the form $X^iH^jY^k$. But $Yu_0 = 0$, and H^ju_0 is just a scalar multiple of u_0. So

$$u_0, u_1 = Xu_0, \quad u_2 = X^2u_0, \cdots$$

span the subspace U. Now u_0 is in $\tilde{L}_0$, u_1 is in $\tilde{L}_1$, etc. So these vectors lie in distinct disjoint subspaces of $\tilde{L}(m, n)$, implying that they are linearly independent. Therefore the string of vectors $u_0, u_1, \ldots$ forms a basis for U. But U is finite dimensional. Call the last vector in the string u_s.

How long is this string of vectors? By the definition of U, none of the three operators X, Y, or H can move anything outside of the subspace U. Thus the restrictions of X, Y, and H to U are operators on U. Denote these operators on U by X', Y', and H'. Note for use below that the relation

$$H' = X'Y' - Y'X'$$

still holds. Now take u_k in $\tilde{L}_k$. Then

$$H'u_k = (2k - mn)u_k,$$

since u_k is a linear combination of the $\tilde{\mathbf{a}}$ with $a_1 + a_2 + \cdots + a_n = k$. So the trace of the operator H' can be easily computed:

$$\operatorname{trace} H' = -mn + (2 - mn) + \cdots + (2s - mn).$$

We've finally arrived at the trick alluded to in the introduction: It's a simple fact that

$$\operatorname{trace} AB = \operatorname{trace} BA$$

for any two linear operators A and B. Apply this fact to the operators X' and Y':

$$\operatorname{trace} X'Y' = \operatorname{trace} Y'X'.$$

Then

$$\operatorname{trace} H' = \operatorname{trace}(X'Y' - Y'X') = 0,$$

implying

$$s = mn.$$

We conclude that the string of vectors

$$u_0, u_1, \ldots, u_{mn}$$

forms a basis for U.

Continue the construction of a new basis by letting β be the smallest index such that $U \cap \tilde{L}_\beta$ is not all of $\tilde{L}_\beta$. Pick any vector v_β in $\tilde{L}_\beta$ which is not in U. Let V denote the subspace spanned by all vectors resulting from all possible compositions of X, Y, and H acting on either u_0 or v_β. Since Yv_β must be in U, one can see that

$$v_\beta, \quad v_{\beta+1} = Xv_\beta, \quad v_{\beta+2} = X^2v, \quad \cdots$$

together with the u_k's span V. If v_q is in U for some index q, then v_r is also in U for all $r \geqslant q$. Let v_t be the last vector in the second string which is not in U. By considering rank subspaces, it's easy to see that the only possible linear dependences among all of the u_k's and v_h's (with $\beta \leqslant h \leqslant t$) must occur between a u_k and a v_k in the same rank subspace $\tilde{L}_k$. But our choice of t prohibits any such degenerate relationships between members of the two strings. The union of the two strings therefore forms a basis for V. The trace trick can be used again to find that

$$t = mn - \beta.$$

So a basis for V is

$$u_0, u_1, \ldots, u_{mn}; \quad v_\beta, v_{\beta+1}, \ldots, v_{mn-\beta}.$$

Repeat this procedure, creating subspaces

$$U \subset V \subset W \subset \cdots \subset \tilde{L}(m, n).$$

Since $\tilde{L}(m, n)$ is finite dimensional, the process must eventually stop. Call the last subspace so constructed Z, and call the last basis vectors so chosen

$$z_\omega, z_{\omega+1}, \ldots, z_{mn-\omega}.$$

All the strings of vectors taken together form a new basis for $\tilde{L}(m, n) = Z$. The subset of these vectors with subscript k forms a basis for the rank subspace $\tilde{L}_k$. Finally note that each string of new basis vectors is symmetric about the middle rank subspace(s): If a string starts in $\tilde{L}_k$ with $k < mn/2$, then it ends in $\tilde{L}_{mn-k}$.

We're now ready to complete the proof of the lemma. If $k < mn/2$, then X_k takes all of the new basis vectors for $\tilde{L}_k$ to new basis vectors for $\tilde{L}_{k+1}$ because no strings end below the middle. So these X_k's are injective, as required. If $k \geqslant mn/2$, then X_k hits every new basis vector for $\tilde{L}_{k+1}$ with a new basis vector from $\tilde{L}_k$, since no strings start above the middle. And so these X_k's are surjective, as required.

We're done! Since this lemma was the last step in our solutions, we can conclude:

THEOREM. *The posets $L(m, n)$ and $M(n)$ are rank unimodal and Sperner. Therefore the proper*

grid shading counts for fixed m and n form a unimodal sequence, and no set of n positive real numbers has more subsets summing to a common sum than does the set $N = \{1, 2, \dots, n\}$.

5. Remarks. Our methods proved the existence of rank matchings for $L(m, n)$ and $M(n)$ satisfying the requirements of the lemma in Section 2, but did not explicitly construct such matchings. Some very nice matchings ("symmetric chain decompositions") have been found for $L(2, n)$, $L(3, n)$, $L(4, n)$, and $L(5, n)$ [Li2], [Rie], [Wes] (and K. Leeb and V. Strehl, unpublished). No rank matchings of any kind (not even ones possibly ignoring order relations) have been found for general $L(m, n)$ and $M(n)$.

The Lie representation constructions employed in this article may seem somewhat *ad hoc* and unnatural with respect to the combinatorial situation at hand. However, it can be shown [Pr1] that a ranked poset is rank unimodal, "rank symmetric," and "strongly Sperner" if and only if it "carries" a representation of $s\mathrm{l}(2, \mathbb{C})$. **Rank symmetric** means $p_k = p_{r-k}$ and **strongly Sperner** means no union of N antichains is bigger than the union of the N largest ranks, for all $N \geq 1$. By **carry**, we mean that three operators X, Y, and H can be defined on the vector space associated with the poset as in this article, except that the operator X need not have all coefficients equal to 1 and the operator Y need not obey the order relations.

The computations required to determine whether a given ranked poset carries a representation of $s\mathrm{l}(2, \mathbb{C})$ can be quite difficult. However, the computations become relatively simple in the context modelled the most closely after the situation in this article: distributive lattices, coefficients all equal to 1 for the operator X, and the operator Y respecting the order relations. Surprisingly, it is possible to prove [Pr3] that this version of our techniques can be applied to only one other infinite family of and two exceptional distributive lattices besides the $L(m, n)$ and $M(n)$. Dynkin-like diagrams play a crucial role in the classification procedure. The set of all of these diagrams also arises as the set of Dynkin-like diagrams corresponding to all quotients of semisimple Lie groups which are Hermitian symmetric spaces!

Stanley was the first to prove that $L(m, n)$ and $M(n)$ have the strong Sperner property [St1]. The lemma in Section 3 is a simpler version of a lemma that he developed for this purpose. To show that the operator X satisfied the requirements of his lemma, Stanley first noted that bases for the cohomology rings of certain projective varieties (the Grassmannians for $L(m, n)$) can be labelled in a natural way with elements of $L(m, n)$ or $M(n)$. Then he observed that multiplication by a hyperplane section in the cohomology ring when viewed as a linear operator is exactly the operator X. Stanley completed his proofs with the hard Lefschetz theorem, a major theorem of algebraic geometry: This theorem states that the linear operator defined by multiplication with a hyperplane section produces a vector space isomorphism between "sister" cohomology groups.

In the cases of $L(m, n)$ and $M(n)$ (Stanley also treated many other posets), it is possible to replace the algebraic geometry in Stanley's proofs with representations of Lie algebras. This was done [Pr2] by combining combinatorial identifications of weights of representations by Hughes [Hug] with the construction of principal three dimensional subalgebras due to Dynkin [Dyn, p. 168] in the case of minuscule representations. This replacement yielded the Lie algebraic proofs from which the proofs given here were derived. The trace trick is a standard technique in representation theory [SaW, p. 278].

When the historical background for this article was being researched, after a little translation it was discovered that the vector spaces and operators used by Cayley [Cay] and Sylvester [Syl] for their problem in invariant theory were nearly the same as those used here for the poset $L(m, n)$. Cayley even used the same letters X, Y, and H! Although Sylvester's proof of the injectivity and the surjectivity of the operators X_i is different, it also uses the commutation relations between the three operators. For two other proofs of the unimodality of the grid shading sequences in the context of classical invariant theory, see Springer [Spr, Ex. 3.3.6(1)] and Elliott [Ell, p. 149].

The unimodality of the grid shading sequences has also been proved using representations of symmetric groups by White [Whi], Towber and Wagner [ToW], and Stanley [St2]. Stanley's proof

is fairly simple, using nothing more sophisticated than general facts about representations of finite groups. He and Larry Harper [Har] have recently extended these methods to show that $L(m, n)$ is strongly Sperner. Their approach turns out to be a stronger version of some work of Pouzet [Pou]. The paper by Towber and Wagner shows that the symmetric group approach and the $sl(2, \mathbb{C})$ approach are closely related.

The original conjecture of Erdős and Moser [Erd] actually concerned *any* $2n$ (or $2n + 1$) real numbers, rather than n positive real numbers. The answer is what one would expect:

$$-n, \quad -n+1, \quad \ldots, \quad -1, \quad 0, \quad 1, \quad \ldots, \quad n-1, \quad (n).$$

This conjecture was also first proved by Stanley [St1]. Roughly speaking, if $n = m + z + p$ specifies how many numbers are negative, zero, and positive, his proof consists of showing $M(m) \times M(z) \times M(p)$ to be Sperner (which can also be done with our methods) together with a short elementary argument showing that $z = 1$ and $n = p(\pm 1)$.

The reduction of the conjecture of Erdős and Moser for positive real numbers to the question of the Spernerity of $M(n)$ is due to Lindström [Li1].

I am indebted to my adviser Richard Stanley for encouragement and mathematical inspiration. I am also indebted to Bruce Sagan and several other people for constructive criticisms.

References

[Cay] A. Cayley, A Second Memoir upon Quantics, Coll. Math. Papers, vol. 2, Cambridge University Press, Cambridge, 1889, pp. 250–275.

[Dyn] E. B. Dynkin, Semisimple subalgebras of semisimple lie algebras, Amer. Math. Soc. Transl., Ser. 2, vol. 6 (1957) 111–244.

[Ell] E. B. Elliott, An Introduction to the Algebra of Quantics, Oxford University Press, London, 1913.

[Erd] P. Erdős, Extremal problems in number theory, Theory of Numbers, A. L. Whiteman, Editor, Amer. Math. Soc. Proc. Symp. Pure Math., vol. 8 (1965) 181–189.

[Hug] J. W. B. Hughes, Lie algebraic proofs of some theorems on partitions, Number Theory and Algebra, H. Zassenhaus, Editor, Academic Press, New York, 1977, 135–155.

[Li1] B. Lindström, Conjecture on a theorem similar to Sperner's, Combinatorial Structures and Their Applications, R. Guy, Editor, Gordon and Breach, New York, 1970, p. 241.

[Li2] ______, A partition of $L(3, n)$ into saturated symmetric chains, European J. Comb., 1 (1980) 61–63.

[Pr1] R. A. Proctor, Representations of $sl(2, \mathbb{C})$ on posets and the Sperner property, SIAM J. Algebraic Discrete Methods, 3(1982)275–280.

[Pr2] ______, Bruhat lattices, plane partition generating functions, and minuscule representations, preprint.

[Pr3] ______, A Dynkin diagram classification of posets satisfying a combinatorial condition, in preparation.

[Rie] W. Riess, Zwei Optimierungsprobleme auf Ordnungen, Arbeitsberichte des Instituts für mathematische Maschinen und Datenverarbeitungen (Informatik) 2 (1978) no. 5 (April) Erlangen.

[SaW] A. Sagle and R. Walde, Introduction to Lie Groups and Lie Algebras, Academic Press, New York, 1973.

[Spr] T. A. Springer, Invariant Theory, Lecture Notes in Mathematics 585, Springer-Verlag, New York, 1977.

[St1] R. P. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, SIAM J. Algebraic Discrete Methods, 1 (1980) 168–184.

[St2] ______, Some aspects of groups acting on finite posets, J. Combin. Theory Ser. A, 32 (1982) 132–161.

[Syl] J. J. Sylvester, Proof of the Hitherto Undemonstrated Fundamental Theorem of Invariants, Coll. Math. Papers, vol. 3, Chelsea, New York, 1973, 117–126.

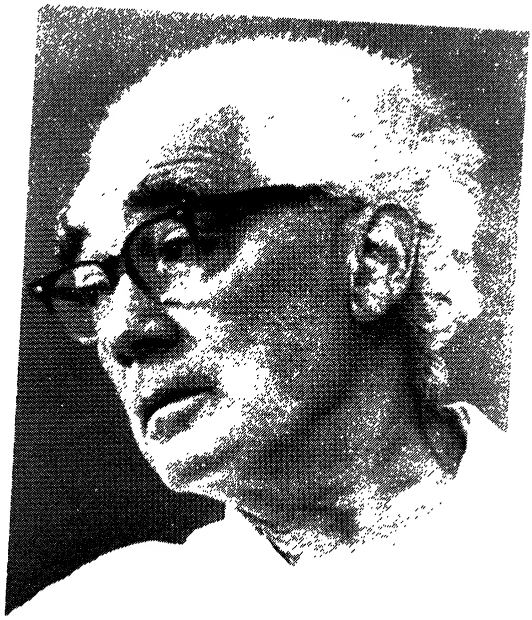
[ToW] J. Towber and B. Wagner, A new proof of a theorem in classical invariant theory, preprint.

[Wes] D. B. West, A symmetric chain decomposition of $L(4, n)$, European J. Combin., 1 (1980) 379–383.

[Whi] D. E. White, Monotonicity and unimodality of the pattern inventory, Advances in Math., 38 (1980) 101–108.

[Har] L. Harper, Morphisms for the strong Sperner property of Stanley and Griggs, preprint.

[Pou] M. Pouzet, Application d'une propriété combinatoire des parties d'un ensemble aux groupes et aux relations, Math. Z., 150 (1976) 117–134.



Probability and set theory, but not necessarily in that order. See page 784.

In the following we use the fact that $|(\sin x)/x| \leqslant 1$. We also need that $\Gamma(x+1)\Gamma(n+1-x)$ is decreasing for $0 < x < n/2$. To show this we use the well-known Beta-function integral

$$\frac{\Gamma(x+1)\Gamma(n+1-x)}{\Gamma(2+n)} = \int_0^1 t^x(1-t)^{n-x}\,dt$$
$$= \tfrac{1}{2}\int_0^1 [t(1-t)]^{n/2}\left\{\left(\frac{t}{1-t}\right)^{(n/2)-x} + \left(\frac{1-t}{t}\right)^{(n/2)-x}\right\} dt.$$

The result follows from the fact that (for $a > 0$) the function $a^x + a^{-x}$ is increasing on $[0, \infty)$, $(a \neq 1)$.

For the sake of simplicity we shall assume that n is even and $k \geqslant 4$; the result is also true for odd n and $k \geqslant 1$. We estimate the integral from 1 to $n/2$ by dividing this into $[1, 3]$ and $[3, n/2]$. We then have

$$\left|\int_1^{n/2}\Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}dx\right| < \frac{2\Gamma(n)}{k-3} + \frac{n}{2}\Gamma(4)\Gamma(n-2)\pi,$$

and hence we find for $n \to \infty$

$$J_{n,k} = \int_0^1 \Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}dx + O(k^{-1}\Gamma(n))$$

uniformly in k for $n \to \infty$. Applying Stirling's formula then yields

$$\frac{J_{n,k}}{n!} = \int_0^1 n^{-x}\left\{1 + O\left(\frac{1}{n}\right)\right\}\frac{\Gamma(x+1)\sin \pi x}{k-x}dx + O\left(\frac{1}{nk}\right).$$

Substitute $t = x \log n$. We find

$$\frac{J_{n,k}}{n!} = \frac{\pi}{k(\log n)^2}\int_0^{\log n} te^{-t}\cdot\Gamma\left(1+\frac{t}{\log n}\right)\cdot\frac{\log n}{\pi t}\sin\frac{\pi t}{\log n}\left(1-\frac{t}{k\log n}\right)^{-1} dt + O\left(\frac{1}{nk}\right).$$

The final step was the following standard argument. Let $f(t) = 1 + O(t)$, $(t \to 0)$ and $0 \leqslant f(t) \leqslant C$ on $[0, 1]$. Then

$$\left|\int_0^m te^{-t}f\left(\frac{t}{m}\right)dt - \int_0^\infty te^{-t}\,dt\right| \leqslant \int_0^{2\log m} te^{-t}O\left(\frac{t}{m}\right)dt + \int_{2\log m}^m te^{-t}O(1)\,dt + \int_m^\infty te^{-t}\,dt$$
$$= O(m^{-1}) + O(m^{-2}\log m) + O(me^{-m}) = O(m^{-1}), \qquad (m \to \infty).$$

Since $f_k(t) = \Gamma(1+t)\cdot\dfrac{\sin \pi t}{\pi t}\cdot\left(1-\dfrac{t}{k}\right)^{-1}$ satisfies these conditions (also for $k = 1$, in fact $f_k(t) \leqslant 2$ on $[0, 1]$ for all k), we have, uniformly in k for $1 \leqslant k \leqslant n-1$,

$$\frac{J_{n,k}}{n!} = \frac{\pi}{k(\log n)^2}\left(1 + O\left(\frac{1}{\log n}\right)\right), \qquad (n \to \infty). \tag{5}$$

The proof for $k = 0$ is similar. The result (2) follows from (4) and (5).

Paul F. Byrd points out that a detailed derivation is given in V. I. Krylov, *Approximate Calculation of Integrals*, The Macmillan Co., pages 84–86.

ANSWERS TO "PHOTOS" ON PAGE 735

Top: William Feller; bottom: Ernst Zermelo.

ON THE EQUATION $xyz = x + y + z = 1$

CHARLES SMALL
Department of Mathematics, Queen's University, Kingston, Ontario, Canada

0. Introduction. Beyond providing results about the specific equation $xyz = x + y + z = 1$, this paper is intended as a palatable introduction to some general techniques in the study of diophantine equations and elliptic curves, particularly over finite fields.

It is known ([C_1], see also [Si_1] and [Si_2]) that the equation of the title has no solutions in the rational field $\mathbb{Q}$. In this paper we discuss solutions in the rings $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m, and in the finite fields $\mathbb{F}_q$ of $q = p^n$ elements, p prime, $n \geqslant 1$. (Recall that the number of elements in a finite field is always a prime power, and conversely for any prime power $q = p^n$ there is, up to isomorphism, exactly one field with q elements, denoted $\mathbb{F}_q$. For $q = p^n$ and $q' = p'^{n'}$ we have $\mathbb{F}_q \subseteq \mathbb{F}_{q'}$ if and only if $p = p'$ and $n|n'$; p is called the *characteristic* of $\mathbb{F}_q$. A good reference for these and other basic facts about finite fields is [IR, chapter 7].)

In $\mathbb{Z}/2\mathbb{Z}$ the equation has the solution $x = y = z = 1$; however, in $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$ there is no solution, as one sees at once from the fact that, in these rings, $xyz = 1$ forces each of x, y, z to be ± 1. (For the same reason, there are no solutions in the ring $\mathbb{Z}$ of integers.) *We show in* §1 *that there are solutions in $\mathbb{Z}/m\mathbb{Z}$ if and only if m is divisible by neither* 3 *nor* 4. The Chinese Remainder Theorem reduces this to the existence of solutions in $\mathbb{Z}/p^n\mathbb{Z}$ for all $n \geqslant 1$ and all primes $p \neq 2, 3$. For fixed $p \neq 2$, it turns out that existence of solutions in $\mathbb{Z}/p^n\mathbb{Z}$ for all n is equivalent to existence of solutions in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$; this follows from a general lifting lemma as discussed in §1 below. Finally, we establish the existence of solutions in $\mathbb{Z}/p\mathbb{Z}$ ($p \neq 2, 3$) directly, by quadratic reciprocity.

In particular, then, the results of §1 provide solutions in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for all $p \neq 3$, and therefore in all fields of characteristic $p \neq 3$. *In* §2 *we give a precise count of the number of solutions in $\mathbb{F}_q$, for arbitrary q.* It turns out, for example, that although there are no solutions in $\mathbb{F}_3$, there are twelve in $\mathbb{F}_9 = \mathbb{F}_3(i)$, $i^2 = -1$, namely the six permutations of $(1, i, -i)$ and the six permutations of $(1 + i, 1 - i, -1)$. To give a further example, we will see that the number of solutions in $\mathbb{F}_q$ for $q = 11^n$, $n = 1, 2, 3, 4, 5$ is 3, 105, 1311, 14685 and 161523 respectively.

The equation under consideration can be read over any field, and except for a few fields over which it becomes *singular*, it defines an *elliptic curve*. The terminology is explained in §2; it turns out that our equation is singular only in characteristics 2 and 13. (Viewing the equation as an elliptic curve is basic to the proof in [C_1] that there are no rational solutions. A "direct" proof is given in [Si_2].) A fundamental theorem on elliptic curves defined over $\mathbb{Q}$ (see [C_2]) states that even when (as in our case) there are no rational solutions, there are always p-adic solutions—that is, solutions modulo p^n for all $n \geqslant 1$—for all primes p outside a finite set which can be determined from the equation. Our result in §1 shows that for the elliptic curve given over $\mathbb{Q}$ by $xyz = x + y + z = 1$ the only primes which have to be excluded are 2 and 3. The proof here makes no reference to elliptic curves and is completely elementary; the only tools required are quadratic reciprocity and the lifting lemma already referred to. In §2 we make the connection with elliptic curves explicit, and invoke rationality of the zeta function to count numbers of solutions in $\mathbb{F}_q$. However, we emphasize that *the reader does not need to know anything about elliptic curves to follow the argument*. On the contrary, it is hoped that the simple treatment here of an explicit example will whet the appetite for deeper study of the general theory, for example via [IR] (especially chapters 10 and 11), [C_2], [H'] (especially chapter IV, §4), [J], [R], or [T].

Charles Small received his Ph.D. under Hyman Bass at Columbia University in 1969 and has been at Queen's University since 1970. He has published more than a dozen papers in algebra and number theory and coauthored two books, one on the Brauer group of a commutative ring and one on homological methods in commutative algebra. He is a violist and the composer of an as-yet-unperformed string quartet, and he avidly collects (and hereby solicits) serendipitous iambic pentameters.

1. Solutions in $\mathbb{Z}/m\mathbb{Z}$. We will show:

1.1 THEOREM. *The equation $xyz = x + y + z = 1$ has solutions in $\mathbb{Z}/m\mathbb{Z}$ if and only if m is divisible by neither* 3 *nor* 4.

Write $m = 2^{a_0}p_1^{a_1} \cdots p_t^{a_t}$ where $a_0 \geqslant 0$, $p_1, \ldots, p_t$ are distinct odd primes, and the exponents $a_1, \ldots, a_t$ are positive integers. To solve $xyz = x + y + z = 1$ in $\mathbb{Z}/m\mathbb{Z}$ is to solve the congruence $xyz \equiv x + y + z \equiv 1 \pmod{m}$, and the Chinese Remainder Theorem ensures that this is equivalent to solving the same congruence modulo 2^{a_0} and modulo $p_i^{a_i}$ for all $i = 1, \ldots, t$. Since we already know there is a solution modulo 2 and none modulo 3 or 4, it is therefore clear that Theorem 1.1 follows from:

1.2 THEOREM. *The equation $xyz = x + y + z = 1$ has solutions in $\mathbb{Z}/p^n\mathbb{Z}$ for all $n \geqslant 1$ and all primes $p \neq 2, 3$.*

We will prove 1.2 by reducing to the case $n = 1$ and then using quadratic reciprocity to show existence of solutions in $\mathbb{Z}/p\mathbb{Z}$. First, consider the following attempt to solve $xyz = x + y + z = 1$: eliminate z to get $xy(1 - x - y) = 1$, and arrange this as a quadratic in y: $xy^2 + (x^2 - x)y + 1 = 0$. The discriminant is $(x^2 - x)^2 - 4x$; call this $\Delta(x)$. Now if the ring R we are working in contains an element x such that

(a) $\Delta(x)$ is a square in R, and
(b) we can divide by $2x$ in R,

then the quadratic formula furnishes y and we have a solution to $xyz = x + y + z = 1$ in R. This proves:

1.3 LEMMA. *Let R be a commutative ring in which* 2 *is invertible,* and define $\Delta : R \to R$ by $\Delta(x) = (x^2 - x)^2 - 4x$. If there is an invertible element $x \in R$ for which $\Delta(x)$ is a square, the equation $xyz = x + y + z = 1$ has a solution in R.*

To apply Lemma 1.3 to $R = \mathbb{Z}/p\mathbb{Z}$ we use quadratic reciprocity. Good references are [IR] chapter 5 and [Se] I §3. For convenience, we recall the facts we will need.

Fix an odd prime p, and for each integer m let $\overline{m}$ be the class of m modulo p, that is, $\overline{m}$ is the image of m under the natural mapping $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. Then m is a *quadratic residue* mod p, and we write $(m/p) = 1$, if $\overline{m}$ is a nonzero square in $\mathbb{F}_p$, and m is a *quadratic nonresidue* mod p, written $(m/p) = -1$, if $\overline{m}$ is not a square in $\mathbb{F}_p$. If $\overline{m} = 0$ (that is, if $p|m$) we put $(m/p) = 0$. The mapping $\left(\frac{-}{p}\right): \mathbb{Z} \to \{0, \pm 1\}$ so defined is called the Legendre symbol mod p. Clearly (m/p) depends only on $\overline{m}$, that is

$$\left(\frac{m + kp}{p}\right) = \left(\frac{m}{p}\right) \text{ for all } k.$$

It is also clear that changing m by a square doesn't change (m/p), that is,

$$\left(\frac{mk^2}{p}\right) = \left(\frac{m}{p}\right) \text{ for all } k.$$

The other basic properties are:

$$\text{(i)}\quad \left(\frac{mr}{p}\right) = \left(\frac{m}{p}\right)\left(\frac{r}{p}\right)$$

*Recall that $s \in R$ is *invertible* if $st = 1$ for some $t \in R$. Thus to require $s \in R$ to be invertible is to require that we may divide by s (= multiply by t) in R. For example, 2 is invertible in $\mathbb{Z}/p^n\mathbb{Z}$ and in $\mathbb{F}_q$, $q = p^n$, for all primes $p \neq 2$ and all $n \geqslant 1$.

(ii) $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \pmod 4$

(iii) $\left(\frac{2}{p}\right) = 1$ if and only if $p \equiv \pm 1 \pmod 8$

and (iv) For an odd prime q, $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$ unless both p and q are $\equiv 3 \pmod 4$, in which case $\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right)$.

The necessary step to apply 1.3 to $\mathbb{F}_p$ is:

1.4 LEMMA. *Let $p > 3$ be prime and define $\Delta: \mathbb{F}_p \to \mathbb{F}_p$ as in* 1.3. *There exists $0 \neq x \in \mathbb{F}_p$ such that $\Delta(x)$ is a square.*

Proof. We will show, more precisely, that $x = 1$ works unless $p \equiv 7, 11, 19$ or $-1 \pmod{24}$; that $x = -1$ works when $p \equiv 7$ or $-1 \pmod{24}$; that $x = 3$ works when $p \equiv 19 \pmod{24}$; and that for each $p \equiv 11 \pmod{24}$ either $x = -3$ or $x = -4$ will work. (Here "$x = 3$" means, of course, "$x = \bar{3} \in \mathbb{F}_p$", etc.)

If $p \equiv 1 \pmod 4$ then $(-1/p) = 1$, and since $\Delta(1) = -1 \cdot 2^2$ we may take $x = 1$. This leaves $p \equiv 7, 11, 19$ and $-1 \pmod{24}$. If $p \equiv 7$ or $-1 \pmod{24}$, then $p \equiv -1 \pmod 8$, so $(2/p) = 1$. Since $\Delta(-1) = 2^3$, we can take $x = -1$ in this case. If $p \equiv 19 \pmod{24}$, then $p \equiv 3 \pmod 8$ and $p \equiv 1 \pmod 3$, so that $(2/p) = -1$ and $(3/p) = -(p/3) = -1$. Hence, since $\Delta(3) = 2^3 \cdot 3$ we can take $x = 3$. Finally, suppose $p \equiv 11 \pmod{24}$. Then $p \equiv 3 \pmod 8$ and $p \equiv -1 \pmod 3$, so that $(2/p) = -1$ and $(3/p) = -(p/3) = 1$. Now observe that $\Delta(-3) = 2^2 \cdot 3 \cdot 13$, and $\Delta(-4) = 2^5 \cdot 13$. If $(13/p) = 1$, then we can use $x = -3$; and if $(13/p) = -1$, we can use $x = -4$. This completes the proof of 1.4.

Combining 1.3 and 1.4, we find:

1.5 PROPOSITION. *The equation $xyz = x + y + z = 1$ has solutions in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ for all primes $p \neq 3$.*

To prove Theorem 1.2 (hence also Theorem 1.1) we want to lift solutions mod p (guaranteed by 1.5) to solutions mod p^n, $n \geqslant 1$. Because of the criterion given by Lemma 1.3, the essential requirement is that we be able to lift squares. Thus we need:

1.6 LEMMA. *Let p be an odd prime and m an integer, not divisible by p. If m is a square* mod p *(that is, if $(m/p) = 1$), then in fact m is a square* mod p^n *for all $n \geqslant 1$.*

Lemma 1.6 is a very special case of a general lifting theorem, so we will not prove it here; see [Se], Cor. 2 of II §2, or [BS], Theorem 1 of Chapter 1, §6. (Note, however, with respect to the first of these references, that the statement of Theorem 1, which Cor. 2 depends upon, should have $0 < j$ and $0 \leqslant 2k$ in place of $0 \leqslant j$ and $0 < 2k$.) For a more elementary treatment see [IR], Prop. 4.2.3. For the algebraically inclined reader we remark that Lemma 1.6 can be rephrased as follows: *an integer m, not divisible by the odd prime p, is a square modulo p if and only if it is a square in the ring $\mathbb{Z}_p$ of p-adic integers.* (Similarly, existence of solutions to $xyz \equiv x + y + z \equiv 1$ (modulo p^n) for all $n \geqslant 1$ is equivalent to existence of solutions to $xyz = x + y + z = 1$ in $\mathbb{Z}_p$.) Either of the references [Se] or [BS] cited above in connection with the proof of 1.6 will provide further information about $\mathbb{Z}_p$ and its relation to the $\mathbb{Z}/p^n\mathbb{Z}$; for algebraically minded readers we remark that $\mathbb{Z}_p$ can be defined quickly as $\varprojlim \mathbb{Z}/p^n\mathbb{Z}$.

We are now in a position to prove Theorem 1.2.

Proof of 1.2: We saw in proving 1.4 that for any prime $p > 3$ we can choose $x \in \{\pm 1, \pm 3, -4\}$

such that $(\Delta(x)/p) = 1$. Since $p > 3$, both x and 2 are invertible in $\mathbb{Z}/p^n\mathbb{Z}$. Finally, $\Delta(x)$ is a square in $\mathbb{Z}/p^n\mathbb{Z}$ by 1.6, so we are done by Lemma 1.3 applied to $R = \mathbb{Z}/p^n\mathbb{Z}$.

A curious by-product of all this is the observation that $(x^2 - x)^2 - 4x$ is never a square for $0 \neq x \in \mathbb{Z}$: if it were, the arguments above would furnish a solution in $\mathbb{Q}$ to $xyz = x + y + z = 1$, contrary to [C_1].

In the next section we will lift in a different sense, from $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ to $\mathbb{F}_q$ ($q = p^n$, $n \geqslant 1$), and our results will yield a precise count of the number of solutions of $xyz = x + y + z = 1$ in $\mathbb{F}_q$. It is certainly natural to ask for a similarly quantitative version of Theorem 1.2: *how many* solutions does $xyz = x + y + z = 1$ have in $\mathbb{Z}/p^n\mathbb{Z}$? There is a theory, analogous to what we shall use in §2 for the finite-field case, which bears on this question (see [H] §7 and [Ig]), but we shall not pursue this here.

2. Solutions in Finite Fields. We saw (1.5) that the equation $xyz = x + y + z = 1$ has solutions in the prime fields $\mathbb{F}_p$ for any prime $p \neq 3$, and hence in any finite field of characteristic $\neq 3$. (In fact, as we shall see, $\mathbb{F}_3$ is the *only* finite field in which there is no solution.) In this section we show how the rationality of the zeta function of an elliptic curve can be exploited to count *how many* solutions the equation has in any finite field of characteristic $\neq 2, 13$; the result is 2.7 below. The two excluded characteristics are treated towards the end of the section (see 2.9).

For fixed p, let $\#_n$ denote the number of solutions $(x, y) \in F^2 = F \times F$ to the equation $xy(1 - x - y) = 1$ (or equivalently the number of solutions $(x, y, z) \in F^3 = F \times F \times F$ to $xyz = x + y + z = 1$), where $F = \mathbb{F}_q$ is the field with $q = p^n$ elements, $n \geqslant 1$.

The first step toward a computation of $\#_n$ is the observation that the argument given in §1 to show $\#_1 > 0$ for $p \neq 3$ can be refined to yield a computation of $\#_1$:

2.1 LEMMA. *For any odd prime p, the number $\#_1$ of solutions to $xyz = x + y + z = 1$ in $\mathbb{F}_p$ is*

$$p - 1 + \sum_{x=1}^{p-1} \left(\frac{\Delta(x)}{p} \right)$$

where $\Delta(x) = (x^2 - x)^2 - 4x$ and where $(\Delta(x)/p)$ is the Legendre symbol defined in §1.

Proof. As in §1, rewrite the equation as a quadratic in y: $xy^2 + (x^2 - x)y + 1 = 0$. The discriminant is $\Delta(x)$. Each nonzero element of $\mathbb{F}_p$ is $\bar{x}$ for some integer $x \in S = \{1, \ldots, p - 1\}$. From the quadratic formula we see that each $x \in S$ with $(\Delta(x)/p) = 0$ contributes exactly one solution; each $x \in S$ with $(\Delta(x)/p) = 1$ contributes two solutions; and each $x \in S$ with $(\Delta(x)/p) = -1$ contributes no solution. Hence each $x \in S$ contributes $(\Delta(x)/p) + 1$ solutions, and the total number of solutions is

$$\sum_{x=1}^{p-1} \left(\left(\frac{\Delta(x)}{p} \right) + 1 \right) = p - 1 + \sum_{x=1}^{p-1} \left(\frac{\Delta(x)}{p} \right).$$

REMARK: It follows from 2.1 that $\#_1$ can be zero only if $(\Delta(x)/p) = -1$ for *every* $x = 1, 2, \ldots, p - 1$. The argument in §1 (via quadratic reciprocity) shows that this never happens for $p > 3$; indeed, we saw that for each $p > 3$, some x in $\{\pm 1, \pm 3, -4\}$ satisfies $(\Delta(x)/p) = 1$. An attractive and important problem is the evaluation of sums of the form $\sum_{x=0}^{p-1}(f(x)/p)$ where f is a polynomial with integer coefficients.

The next step towards a computation of $\#_n$ is to "homogenize" the polynomial

$$f(x, y) = xy(1 - x - y) - 1 = xy - x^2y - xy^2 - 1;$$

that is, we consider the *homogeneous* polynomial

$$\bar{f}(x, y, z) = xyz - x^2y - xy^2 - z^3$$

obtained from f by multiplying each term by the appropriate power of the new variable z. Clearly the number $\#_n$ of zeroes of f in $F^2 = F \times F$ is the same as the number of zeroes of $\bar{f}$ in

$F^3 = F \times F \times F$, *if we count the latter properly*. Indeed, if we identify *proportional* solutions to the equation $\bar{f} = 0$ (count two solutions (x_1, y_1, z_1) and (x_2, y_2, z_2) as the *same* solution if $x_2 = \lambda x_1, y_2 = \lambda y_1, z_2 = \lambda z_1$, for some $0 \neq \lambda \in F$) then $\#_n$ is exactly the number of solutions (x, y, z) to $\bar{f} = 0$ with $z \neq 0$.

More formally, let N_n be the number of equivalence classes of solutions in F^3 to $\bar{f} = 0$, where the equivalence relation is proportionality, as above, and where the trivial solution $(x, y, z) = (0,0,0)$ is not counted. Then, to use the traditional geometric language, N_n is the number of points on the *projective* curve over F given by the equation $\bar{f} = 0$; the points with $z = 0$ are "at infinity," and those with $z \neq 0$ are said to be "at finite distance," or to lie on the *affine* curve over F given by $f = 0$.

One may well ask, why worry about the number N_n of points on the projective curve $\bar{f} = 0$ when what we are interested in is the number $\#_n$ of points on the affine curve $f = 0$. One answer is that some powerful theorems are available for computing N_n, and the number $N_n - \#_n$ of points at infinity is easily computed. We return to this question at the end of §3.

How many points at infinity does our curve $f = 0, f(x, y) = xy - x^2y - xy^2 - 1$, have? They are the solutions to $xyz - x^2y - xy^2 - z^3 = 0$ with $z = 0$, that is, the solutions to $xy(x + y) = 0$ (where "solution" means "proportionality class of solutions, with $x = y = z = 0$ excluded"). Clearly there are exactly three: $x = 0$, $y = 0$, and $x = -y$; that is, $(x, y, z) = (0,1,0)$, $(1,0,0)$, and $(1, -1, 0)$. Thus:

2.2 Lemma. $\#_n = N_n - 3$. *In particular,*

$$N_1 = p + 2 + \sum_{x=1}^{n-1} \left(\frac{\Delta(x)}{p} \right),$$

where $\Delta(x) = (x^2 - x)^2 - 4x$.

Our next task is to describe the theory used to compute N_n. For this we start with *any* polynomial f in, say, m variables, with coefficients in $\mathbb{F}_p$. We will define the zeta function $Z(t)$ of f, state the theorem which computes $Z(t)$ in the case when f defines an elliptic curve, and show how this applies to the polynomial $f(x, y) = xy(1 - x - y) - 1$.

Let $\bar{f}$ be the "homogenization" of f, obtained as above by multiplying each monomial in f by a suitable power of a new variable; thus $\bar{f}$ is a homogeneous polynomial in $m + 1$ variables over $\mathbb{F}_p$. Let N_n be the number of solutions in F^{m+1} to $\bar{f} = 0$, where $F = \mathbb{F}_q$, $q = p^n$. (Again, "solution" in this context means "proportionality class of solutions, with $(x_1, x_2, \ldots, x_{m+1}) = (0,0,\ldots,0)$ excluded.") Then the *zeta function* of f is, by definition, the power series

$$Z(t) = \exp \sum_{n \geq 1} N_n \frac{t^n}{n}.$$

It is simplest, and quite sufficient for our purposes, to consider $Z(t)$ as a formal power series in t with rational coefficients. In other words, $Z(t)$ is obtained by formal substitution of $x = \Sigma_{n \geq 1} N_n t^n / n$ in $\exp(x) = \Sigma_{i \geq 0} x^i / i!$. Since the power series x has zero constant term, the result is a well-defined power series over $\mathbb{Q}$.

We are now ready to specialize to the case of elliptic curves. If $\bar{f}$ is homogeneous of degree 3, in three variables, one says $\bar{f}$ is a *projective plane cubic*; if in addition $\bar{f}$ is *nonsingular*, one says f (or $\bar{f}$) defines an *elliptic curve* over $\mathbb{F}_p$. The condition for nonsingularity of $\bar{f}$ is that no point $(x, y, z) \in F^3$ should satisfy:

$$(x, y, z) \neq (0,0,0), \bar{f}(x, y, z) = \bar{f}_x(x, y, z) = \bar{f}_y(x, y, z) = \bar{f}_z(x, y, z) = 0,$$

where F is the algebraic closure of $\mathbb{F}_p$ and $\bar{f}_x, \bar{f}_y, \bar{f}_z$ are the partial derivatives of $\bar{f}$ with respect to x, y, z respectively. (This is the "Jacobian criterion" for nonsingularity; the requirement is simply that at no point on the curve do all the partial derivatives vanish; intuitively, the meaning is that at each point on the curve there is a well-defined tangent direction.)

In our example $f(x, y) = xy(1 - x - y) - 1$ this is all particularly easy: we have $\bar{f} = xyz - x^2y - xy^2 - z^3$ so that

$$\bar{f}_x = y(z - 2x - y), \bar{f}_y = x(z - x - 2y), \bar{f}_z = xy - 3z^2.$$

Suppose $P = (x, y, z)$ is a singular point on the curve (that is, a point where $\bar{f}, \bar{f}_x, \bar{f}_y, \bar{f}_z$ all vanish). P cannot be one of the points at infinity, for if $z = 0$ the vanishing of $\bar{f}_z$ shows $x = 0$ or $y = 0$, and the remaining conditions give $(x, y, z) = (0,0,0)$. Once we know $z \neq 0$ we may assume $z = 1$, and the conditions $\bar{f}_x(x, y, z) = \bar{f}_y(x, y, z) = 0$ become

$$(*)\begin{cases} 2x + y = 1 \\ x + 2y = 1 \end{cases}$$

If $p = 3$, it's clear that (*) has no solution so f is nonsingular in this case. For any $p \neq 3$, (*) has the unique solution $x = y = 1/3$, and either of the remaining conditions $\bar{f} = 0$ or $\bar{f}_z = 0$ shows that we must have $p = 2$ or $p = 13$. Thus:

2.3 PROPOSITION. *The equation $xy(1 - x - y) = 1$ defines an elliptic curve over any field of characteristic $p \neq 2, 13$. When $p = 2, (x, y) = (1, 1)$ is the only singular point, and when $p = 13, (x, y) = (9, 9)$ is the only singular point. (In either case, the singular point is $x = y = 1/3$.)*

The remarkable theorem about the zeta function of an elliptic curve is:

2.4 THEOREM. *The zeta function $Z(t)$ of an elliptic curve over $\mathbb{F}_p$, as defined above, is a rational function of t. More precisely, put $c = N_1 - p - 1$, then*

$$Z(t) = \frac{1 + ct + pt^2}{(1 - t)(1 - pt)} = 1 + \frac{N_1 t}{1 - (p + 1)t + pt^2}.$$

For our curve $xy(1 - x - y) = 1$ we have already seen how p determines c:

2.5 LEMMA. *Let $\Delta(x) = (x^2 - x)^2 - 4x$, then $c = 1 + \sum_{x=1}^{p-1} \left(\frac{\Delta(x)}{p}\right)$.*

Proof. Compare the definition of c in 2.4 with the computation of N_1 in 2.2.

For proof and discussion of the fundamental result 2.4 the reader is referred to [J], IX especially §3, and [R], IV especially §2; see also [I-R], XI. It should be noted that this result, powerful as it is, is actually just the tip of a large iceberg. We have defined the zeta function $Z(t)$ only for a curve (i.e., a single polynomial f of two variables), but it can be defined for *any* variety over a finite field. The analogue of 2.4 is true in the general case: the zeta function of *any* variety is a rational function, of a fairly specific form. Of course this explicit form is more complicated in higher dimensions than in the case of a curve,—indeed, it is more complicated for curves in general than for elliptic curves. However, even in the most general case, as we shall see is true in the elliptic curve case, the zeta function encodes, in an extremely efficient way, all the information about N_n as a complicated but predictable function of n.

The point of Theorem 2.4 is that it allows us to use our knowledge of c (equivalently, N_1) to compute N_n for *all* n. To see this, let β be a (complex) root of the numerator $1 + ct + pt^2$ of $Z(t)$ and put $\alpha = 1/\beta$, then

$$1 + ct + pt^2 = (1 - \alpha t)\left(1 - \frac{p}{\alpha}t\right).$$

In terms of α (which is determined by c, hence by p (2.5)), the N_n are given by:

2.6 THEOREM. *For $p \neq 2, 13$ and for all $n \geq 1$, we have $N_n = p^n + 1 - \alpha^n - \left(\frac{p}{\alpha}\right)^n$.*

Note that $-\alpha - (p/\alpha) = c$, so that the special case $n = 1$ of 2.6 is just the definition of c. Putting 2.6 together with 2.2 yields:

2.7 COROLLARY. *The number $\#_n$ of solutions to $xyz = x + y + z = 1$ in $\mathbb{F}_q$, $q = p^n$, $p \neq 2, 13$, is given by $\#_n = p^n - 2 - \alpha - (p/\alpha)^n$ with α determined from p as above*:

$$c = 1 + \sum_{x=1}^{p-1}\left(\frac{\Delta(x)}{p}\right), 1 + ct + pt^2 = (1 - \alpha t)\left(1 - \frac{p}{\alpha}t\right).$$

Thus, once 2.6 is proved, we have in 2.7 an explicit computation of $\#_n$.

Proof of 2.6. From 2.4 we have

$$\frac{(1 - \alpha t)\left(1 - \frac{p}{\alpha}t\right)}{(1 - t)(1 - pt)} = \exp \sum_{n \geqslant 1} N_n \frac{t^n}{n},$$

hence

$$\sum_{n \geqslant 1} N_n \frac{t^n}{n} = \log(1 - \alpha t) + \log\left(1 - \frac{p}{\alpha}t\right) - \log(1 - t) - \log(1 - pt).$$

Since

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$

we have $-\log(1 - x) = \sum_{n \geqslant 1} x^n/n$. Applying this to $x = \alpha t$, $(p/\alpha)t$, t, and pt, we have

$$\sum_{n \geqslant 1} N_n \frac{t^n}{n} = -\sum_{n \geqslant 1} \frac{(\alpha t)^n}{n} - \sum_{n \geqslant 1} \frac{\left(\frac{p}{\alpha}t\right)^n}{n} + \sum_{n \geqslant 1} \frac{t^n}{n} + \sum_{n \geqslant 1} \frac{(pt)^n}{n} = \sum_{n \geqslant 1}\left(p^n + 1 - \alpha^n - \left(\frac{p}{\alpha}\right)^n\right)\frac{t^n}{n}$$

which gives the result upon comparing coefficients of like terms.

A parenthetical observation: if we differentiate the relation

$$\sum_{n \geqslant 1} N_n \frac{t^n}{n} = \log \frac{1 + ct + pt^2}{(1 - t)(1 - pt)}$$

on both sides with respect to t, we find that the power series $\sum_{n \geqslant 1} N_n t^{n-1}$ is a rational function in t. By an ancient and well-known theorem (see for example [S]) a power series is a rational function if and only if its coefficients eventually satisfy a linear recurrence relation. Thus the N_n, and hence also the $\#_n$, satisfy a linear recurrence relation, which can be found explicitly by computing the rational function

$$\frac{d}{dt} \log \frac{1 + ct + pt^2}{(1 - t)(1 - pt)}.$$

We leave the details to the reader.

We want next to illustrate Theorem 2.6 and its Corollary 2.7 with a few numerical examples. Here is a table giving values of c for the first twenty-five primes (with thanks to Wendy Vincent for the computations):

p	3	5	7	11	13	17	19	23	29	31	37	41
c	-1	3	1	-6	-1	3	-2	0	-6	4	7	0

p	43	47	53	59	61	67	71	73	79	83	89	97	101
c	1	-3	0	6	-8	-14	3	-2	-8	-12	6	10	12

To indicate how this information determines the $\#_n$ for each of these primes, consider first $p = 3$. We have $c = -1$, so to determine α we use

$$1 - t + 3t^2 = (1 - \alpha t)\left(1 - \frac{3}{\alpha}t\right).$$

Thus $\alpha = (1 + \sqrt{-11})/2$ will do. Using

$$N_1 = 3 + 1 - \alpha - \left(\frac{3}{\alpha}\right)$$

and

$$N_2 = 9 + 1 - \alpha^2 - \left(\frac{3}{\alpha}\right)^2$$

(from 2.6) we find $N_1 = 3$ and $N_2 = 15$. Remembering to discount the three solutions at infinity, we get $\#_1 = 0$ (as we already knew: $xyz = x + y + z = 1$ has no solutions in $\mathbb{F}_3$) and $\#_2 = 12$ (as indicated in the Introduction).

For $p = 11$ we have $c = -6$, and from

$$1 - 6t + 11t^2 = (1 - \alpha t)\left(1 - \frac{11}{\alpha}t\right)$$

we find that $\alpha = 3 + i\sqrt{2}$ will do. Then

$$\alpha^2 = 7 + 6i\sqrt{2},\ \alpha^3 = 9 + 25i\sqrt{2},\ \alpha^4 = -23 + 84i\sqrt{2},\ \alpha^5 = -237 + 229i\sqrt{2},$$

and using

$$N_n = 11^n + 1 - \alpha^n - \left(\frac{11}{\alpha}\right)^n$$

yields

$$N_1 = 6,\ N_2 = 108,\ N_3 = 1314,\ N_4 = 14688,\ N_5 = 161526.$$

Again remembering to subtract 3 from each of these numbers to allow for the points at infinity, we find the numbers $\#_1, \ldots, \#_5$ indicated in the Introduction. The three solutions in the prime field $\mathbb{F}_{11}$ are the permutations of $(x, y, z) = (7, 7, 9) = (-4, -4, -2)$.

The reader will have noticed that in both of these examples ($p = 3, 11$) we have $p/\alpha = \bar{\alpha}$, the complex conjugate of α. This is, of course, no accident! Indeed, from the equations

$$(1 - \alpha t)\left(1 - \frac{p}{\alpha}t\right) = 1 + ct + pt^2 = (1 - \bar{\alpha}t)\left(1 - \frac{p}{\bar{\alpha}}t\right)$$

it is easy to see that $p/\alpha = \bar{\alpha} \Leftrightarrow \alpha \notin \mathbb{R}$ (the reals). In fact:

2.8 PROPOSITION. *The following are equivalent*:

(i) $|\alpha| = \sqrt{p}$
(ii) $\dfrac{p}{\alpha} = \bar{\alpha}$
(iii) $\alpha \notin \mathbb{R}$
(iv) $c^2 - 4p < 0$
(v) $1 + ct + pt^2$ *has no real root.*
(vi) $|c| < 2\sqrt{p}$
(vii) $|N_1 - p - 1| < 2\sqrt{p}$

The equivalence of (iii) through (vi) results from the fact that $c^2 - 4p$ is the discriminant of the quadratic polynomial $1 + ct + pt^2$ of which $1/\alpha$ is a root, and (vi) $\Leftrightarrow$ (vii) comes from the definition of c (2.4).

A celebrated theorem due to Hasse (1936) states that the conditions listed in Prop. 2.8, besides being equivalent, all hold true. It then follows from 2.6, since

$$N_n = p^n + 1 - \alpha^n - \left(\frac{p}{\alpha}\right)^n = p^n + 1 - \alpha^n - \bar{\alpha}^n,$$

that

$$|N_n - p^n - 1| \leqslant |\alpha|^n + |\bar{\alpha}|^n = 2p^{n/2}.$$

Thus to the list of conditions in 2.8 we can add:

(viii) $|N_n - p^n - 1| \leqslant 2p^{n/2}$ *for all* $n \geqslant 1$, *with strict inequality for* $n = 1$.

The term "Riemann hypothesis" is used in the literature both for statement (i), $|\alpha| = \sqrt{p}$, and for the equivalent statement (viii) that

$$|N_n - p^n - 1| \leqslant 2p^{n/2}.$$

In the latter form it is a strong estimate for the number of solutions: N_n is within $2p^{n/2}$ of $p^n + 1$.

In terms of the number $\#_n = N_n - 3$ of affine solutions to our equation $xy(1 - x - y) = 1$, (viii) can be written as follows:

$$|\#_n - p^n + 2| \leqslant 2p^{n/2}.$$

One trivial consequence of this is that for $p = 3$ and $n > 1$, $\#_n$ is not 0. With 1.5, this justifies the statement that $\mathbb{F}_3$ is the only finite field in which the equation $xyz = x + y + z = 1$ has no solution.

Hasse's theorem was generalized in 1947 by Weil to the case of arbitrary curves, and formed the basis for one of the famous "Weil conjectures," settled recently by Deligne. The fact in question,—in our case, the fact that $|\alpha| = \sqrt{p}$,—is known as the Riemann hypothesis (for varieties defined over finite fields). The classical Riemann hypothesis, of course, is the assertion (as yet unproved) that every complex zero of the Riemann zeta function

$$\zeta(s) = \sum_{n \geqslant 1} n^{-s}$$

satisfies $\mathrm{Re}(s) = 1/2$. To begin to see the connection between these two (one a theorem and one a conjecture, and both referred to as "Riemann hypothesis"), observe that the function $\mathfrak{z}$ of a complex variable s defined by

$$\mathfrak{z}(s) = Z(p^{-s}) = \frac{(p^s - \alpha)(p^s - \bar{\alpha})}{(p^s - 1)(p^s - p)}$$

has all its zeroes on the line $\mathrm{Re}(s) = 1/2$, for if $s = \sigma + it$ and $\mathfrak{z}(s) = 0$ we have

$$\sqrt{p} = |\alpha| = |p^s| = p^\sigma$$

so that $\sigma = 1/2$. It turns out that $\mathfrak{z}(s)$ can be expressed as

$$\mathfrak{z}(s) = \sum_{\underline{d} > 0} N(\underline{d})^{-s}$$

where the sum is over all "positive divisors" $\underline{d}$ on the curve, and for each $\underline{d}$, $N(\underline{d})$ (the "norm" of $\underline{d}$) is a certain power of p; the analogy between $\mathfrak{z}(s)$ and the classical Riemann zeta function is then clear.

For $p = 13$ the general machinery described above doesn't apply for our curve $xy(1 - x - y) = 1$, since the equation becomes singular in characteristics 2 and 13. However, the value $c = -1$ (computed as in 2.5) yields the correct value $N_1 = 13$; the ten (affine) solutions to $xyz = x + y + z = 1$ in $\mathbb{F}_{13}$ are (9, 9, 9), three permutations of (2, 2, 10), and six of (1, 5, 8). For $n > 1$, the formula

$$N_n = 13^n + 1 - \alpha^n - \left(\frac{13}{\alpha}\right)^n$$

breaks down; for example it predicts $N_2 = 195$ whereas in fact $N_2 = 169$. Because our curve is singular in characteristic 2 and 13, additional techniques are needed to compute N_n in these two cases. For the sake of completeness we give the results, though without proof:*

2.9 THEOREM. *For $p = 2$ we have $N_n = 2^n$ if n is even, $2^n + 2$ if n is odd. For $p = 13$ we have $N_n = 13^n$. Thus the number $\#_n$ of solutions to $xyz = x + y + z = 1$ in $\mathbb{F}_q$, $q = p^n$, is $2^n - 3$ when $p = 2$ and n is even, $2^n - 1$ when $p = 2$ and n is odd, and $13^n - 3$ when $p = 13$.*

To close this section we list a few additional facts, leaving proofs as exercises:

2.10 PROPOSITION. *Suppose the elliptic curve E and the prime p are such that $c = 0$. (Since $-c = \alpha + \bar{\alpha}$, $c = 0$ is equivalent to $\mathrm{Re}(\alpha) = 0$; for our curve given by $xy(1 - x - y) = 1$ we have seen that it happens, for $p \leqslant 101$, exactly for $p = 23$, 41 and 53.) Then $N_n = p^n + 1 + \varepsilon$ where $\varepsilon = 0$ if n is odd and $\varepsilon = (-1)^{m+1}2p^m$ if $n = 2m$ is even.*

2.11 PROPOSITION. *In general $N_n \equiv 1 - (-c)^n \pmod p$. Equivalently, $(\alpha + \bar{\alpha})^n \equiv \alpha^n + \bar{\alpha}^n \pmod p$.*

3. Normal Forms. We have, following [R], defined an elliptic curve as an absolutely nonsingular projective plane cubic curve. Elliptic curves have a rich history, and are studied in the literature at various levels of generality; for an indication of how extensive the literature is, see the bibliographies of [C_2] and [T]. More general definitions, for example, which we will not pursue here, are that an elliptic curve is a curve of genus 1, or a certain kind of group scheme, or an abelian variety of dimension 1 (see [H′], IV §4; also [T]). An alternate (and more down to earth) definition [Sch] which we do want to pursue here briefly, in connection with our curve $xyz = x + y + z = 1$, is that an elliptic curve over a field F is given by an equation of the form $y^2 = f(x)$ where $f(x)$ is a polynomial of degree 3 or 4 with coefficients in F and with no repeated roots in an algebraic closure of F. (In this approach, equations $y^2 = f(x)$ with $f(x)$ of degree > 4 are "hyperelliptic curves.")

Notice that if we homogenize $y^2 = f(x)$ (where f has degree $\geqslant 3$) and then set the new variable equal to 0, we find $x = 0$, so that any curve of this form has exactly one point at infinity, namely $(x, y, z) = (0, 1, 0)$.

We have almost seen in §2 that our curve $xyz = x + y + z = 1$ is equivalent (via a suitable change of variables) to one of this type. Indeed, if $xy(1 - x - y) = 1$ and we let

$$y' = 2xy + x^2 - x$$

then

$$y'^2 = \Delta(x) = x^4 - 2x^3 + x^2 - 4x.$$

(This can, of course, easily be verified; to see where it comes from, arrange the equation (as before) as a quadratic in y,

$$xy^2 + (x^2 - x)y + 1 = 0,$$

use the quadratic formula to write

$$y = \left(-x^2 + x \pm \sqrt{\Delta(x)}\right)/2x,$$

and solve for $\Delta(x)$.) Conversely, if $y'^2 = \Delta(x)$ and $x \neq 0$, then putting

$$y = (y' + x - x^2)/2x$$

gives $xy(1 - x - y) = 1$. The two recipes are mutual inverses, and it follows (in all characteristics $\neq 2$) that solutions to $xy(1 - x - y) = 1$ are in one-to-one correspondence with solutions other than $(0,0)$ to $y'^2 = \Delta(x)$.

*I thank Alain Robert and Leslie Roberts for some guidance here.

Is this new equation $y'^2 = \Delta(x)$ an elliptic curve in Schmidt's sense? To see this we must check whether $\Delta(x)$ can have repeated roots in an algebraic closure of $\mathbb{F}_p$; and we can expect that the answer will depend on p. Now the condition for a polynomial to have distinct roots is that it and its derivative have no nontrivial common factor, or, equivalently, that its discriminant be nonzero; and (provided the characteristic is not 2 or 3) the discriminant of a fourth degree polynomial

$$x^4 + b_1x^3 + b_2x^2 + b_3x + b_4$$

is $I^3 - 27J^2$ where

$$I = b_4 - \frac{b_1b_3}{4} + \frac{b_2^2}{12} \quad \text{and} \quad J = \frac{b_2b_4}{6} + \frac{b_1b_2b_3}{48} - \frac{b_3^2}{16} - \frac{b_1^2b_4}{16} - \frac{b_2^3}{216}$$

(see [B], II p. 84 and I p. 121). For our polynomial $\Delta(x)$ we have

$$b_1 = -2, b_2 = 1, b_3 = -4, b_4 = 0,$$

which gives (incredibly!) $I = -23/12, J = -181/216$, and therefore $I^3 - 27J^2 = -26$. Hence, except in characteristics 2, 3 and 13, the equation

$$y^2 = \Delta(x) = x^4 - 2x^3 + x^2 - 4x$$

defines an elliptic curve in Schmidt's sense as well.

Note that the number of points on the projective curve obtained by homogenizing $y'^2 = \Delta(x)$, being one more than the number of points on the affine curve $y'^2 = \Delta(x)$, is *two* more than the number of points on the affine curve $xy(1 - x - y) = 1$, and is therefore *one less* than the number of points on the projective curve obtained by homogenizing $xy(1 - x - y) = 1$ (cf. 2.2). Why do these two projective curves, one corresponding to $xy(1 - x - y) = 1$ and the other to $y'^2 = \Delta(x)$, have *different* numbers of points, even though, on the affine level, the simple change of variables

$$(x', y') = (x, 2xy + x^2 - x)$$

takes us from one to the other?

The answer is that whereas the homogenization $xy^2 + x^2y - xyz + z^3$ of $xy(1 - x - y) - 1$ is nonsingular (except in characteristics 2 and 13; see 2.3), the homogenization

$$\bar{f} = y^2z^2 - x^4 + 2x^3z - x^2z^2 + 4xz^3$$

of $y^2 - x^4 + 2x^3 - x^2 + 4x$ is *always* singular. Indeed, the partial derivatives $\bar{f}_x$, $\bar{f}_y$ and $\bar{f}_z$ of $\bar{f}$ are

$$-4x^3 + 6x^2z - 2xz^2 + 4z^3, 2yz^2, \text{and } 2y^2z + 2x^3 - 2x^2z + 12xz^2,$$

respectively, and the point (0, 1, 0) at infinity obviously satisfies $\bar{f} = \bar{f}_x = \bar{f}_y = \bar{f}_z = 0$. (It is the *only* singular point: at any other point we may take $z = 1$; the vanishing of $\bar{f}_y$ gives $y = 0$, and then the simultaneous vanishing of $\bar{f}$ and $\bar{f}_x$ shows that $\Delta(x)$ has a repeated root, which we have seen is impossible, at least in characteristics $\neq 2, 3, 13$.) In the language of algebraic geometry, the two projective curves in question are *birationally equivalent* but not isomorphic; the singular point at infinity on one of them corresponds to two *distinct* points on the other. Although $y'^2 = \Delta(x)$ defines an elliptic curve in the sense of [Sch] (in characteristics $\neq 2, 3, 13$) the corresponding projective curve, being singular, is *not* an elliptic curve in the sense of [R]. It is, however, birationally equivalent to the elliptic curve given by $xy(1 - x - y) = 1$. If we want to count the number $\#_n'$ of points on $y'^2 = \Delta(x)$ over $\mathbb{F}_{p^n}$ (or the number M_n of points on the corresponding projective curve) we cannot apply the machinery of §2 directly (but we do have

$$\#_n' + 1 = M_n = N_n - 1$$

with N_n as in §2).

Finally, since many discussions of elliptic curves, particularly the classical theory over the field of complex numbers, lay heavy emphasis on the Weierstrass normal form $y^2 = x^3 + ax + b$, we want to indicate briefly how our curve looks in this disguise as well.

Given x, y, z such that $xyz = x + y + z = 1$, put

$$x' = (z - 12)/3z, \; y' = 4(x - y)/z;$$

then

$$y'^2 = x'^3 + \frac{23}{3}x' + \frac{362}{27}.$$

Conversely, given x', y' satisfying this latter equation, and with $x' \neq 1/3$, put

$$x = (3x' - 3y' + 11)/(6x' - 2), \; y = (3x' + 3y' + 11)/(6x' - 2), \text{ and } z = 12/(1 - 3x');$$

then $xyz = x + y + z = 1$. These two recipes make sense in any field F of characteristic other than 2 or 3, and they are mutual inverses, giving a one-to-one correspondence between solutions $(x, y, z) \in F^3$ to $xyz = x + y + z = 1$ and solutions $(x', y') \in F^2$, other than $(x', y') = (1/3, \pm 4)$, to

$$y'^2 = x'^3 + \frac{23}{3}x' + \frac{362}{27}.$$

The latter equation defines an elliptic curve in Schmidt's sense provided

$$h(x) = x^3 + \frac{23}{3}x + \frac{362}{27}$$

has no repeated roots. Again, the criterion is that the discriminant not vanish, and for a cubic polynomial $x^3 + ax + b$ the discriminant is $4a^3 + 27b^2$ ([B], II p. 84 and I pp. 72, 83). Putting $a = 23/3$ and $b = 362/27$ we find

$$4a^3 + 27b^2 = 2^9 \cdot 13,$$

so that again 2, 3 and 13 are the only bad characteristics.

Note that the number of points on the projective curve obtained by homogenizing

$$y'^2 = h(x') = x'^3 + \frac{23}{3}x' + \frac{362}{27},$$

being one more than the number of points at finite distance, is *three* more than the number of points on the affine curve $xy(1 - x - y) = 1$, and hence is *equal* to the number of points on the projective curve obtained by homogenizing $xy(1 - x - y) = 1$. The projective curve given by the homogenizations of $y'^2 = h(x')$ is nonsingular in characteristics $\neq 2, 3, 13$: using the Jacobian criterion as above, we find first that the point $(0, 1, 0)$ at infinity is not a singular point, and then as in the previous case that a singular point at finite distance would provide a common root for $h(x')$ and its derivative, which we know is impossible in characteristics $\neq 2$, 3 and 13. In fact the projective curves defined by (the homogenizations of) $xy(1 - x - y) = 1$ and

$$y'^2 = x'^3 + \frac{23}{3}x' + \frac{362}{27}$$

are *isomorphic* rather than just birationally equivalent (leaving aside, as always, characteristics 2, 3 and 13); indeed there is a general theorem which states that for *nonsingular* curves, birational equivalence implies isomorphism. Thus it is no surprise that the two projective curves have the same number of points (in any field of characteristic $\neq 2, 3, 13$).

Since $y'^2 = h(x')$ is nonsingular (in characteristics $\neq 2, 3, 13$) the machinery of §2 *can* be applied directly to count the number $\#''_n$ of solutions in $\mathbb{F}_{p^n}$ (and the number $N_n = \#''_n + 1$ of points on the corresponding projective curve). Indeed, the same argument as in 2.1 shows that the number $\#''_1$ of solutions in the prime field $\mathbb{F}_p$ is

$$p+\sum_{x=0}^{p-1}\left(\frac{h(x)}{p}\right),$$

and therefore the number $\#_1$ of solutions in $\mathbb{F}_p$ to $xyz = x + y + z = 1$ is

$$p-2+\sum_{x=0}^{p-1}\left(\frac{h(x)}{p}\right)$$

and the number N_1 of points on the corresponding *projective* curve over $\mathbb{F}_p$ is

$$p+1+\sum_{x=0}^{p-1}\left(\frac{h(x)}{p}\right).$$

Thus for $p \neq 2, 3, 13$, the number $c = N_1 - p - 1$ of §2 can be computed from the Weierstrass normal form as

$$c=\sum_{x=0}^{p-1}\left(\frac{h(x)}{p}\right),$$

where

$$h(x) = x^3 + \frac{23}{3}x + \frac{362}{27}.$$

The N_n (hence also the $\#_n'' = N_n - 1$ and the $\#_n = \#_n'' - 2 = N_n - 3$ which were the original goal) are then given, in terms of c, by 2.6.

Comparison of these computations for the three curves:

$$(1) \qquad xy(1-x-y) = 1,$$

$$(2) \qquad y^2 = \Delta(x) = x^4 - 2x^3 + x^2 - 4x,$$

and

$$(3) \qquad y^2 = h(x) = x^3 + \frac{23}{3}x + \frac{362}{27},$$

reveals some important general considerations. A simple change of variables, such as the one which took us from (1) to (2), can introduce a singularity and thereby fail to be an isomorphism, and in particular can alter not only the number of affine solutions but also the number of points on the corresponding projective curve. Even a change of variables such as the one which took us from (1) to (3), which preserves nonsingularity and is therefore an isomorphism, can alter the number of affine solutions: by varying the equation we can put more, or fewer, points at infinity, and what is invariant (under isomorphism) is the number of points on the *complete* (i.e., projective) curve.

This may help explain why the number N_n of points on the projective curve is the more natural concept, and why it is N_n (rather than, say, $\#_n$) which appears in the zeta function and plays a role in the general theory. Even if one's ultimate goal is to count numbers of solutions to diophantine equations,—that is, numbers of points on *affine* curves—over finite fields, it is essential to pass to the *projective* curve to get a proper picture of what is going on. The reader wishing to go further in this direction is referred to [F] and [O], which provide an introduction to the theory of algebraic curves (and, thereby, an entrée into the world of algebraic geometry), as well as the works [C_2], [H′], [J], [R] and [T] already cited.

References

[BS] Z. I. Borevich and I. R. Shafarevitch, Number Theory, Academic Press, New York, 1973.

[B] W. S. Burnside and A. W. Panton, The Theory of Equations, 7th ed., vol 1: 1912, vol. 2: 1928, Dover, New York, 1960.

[C_1] J. W. S. Cassels, On a diophantine equation, Acta Arith., 6 (1960) 47–52.

[C_2] ______, Diophantine equations with special reference to elliptic curves, J. London Math. Soc., 41 (1966) 193–291.

[F] W. Fulton, Algebraic Curves, Benjamin, New York, 1969.

[H] S. J. Haris, Number theoretical developments arising from the Siegel formula, Bull Amer. Math. Soc., 2 (1980) 417–433.

[H′] R. Hartshorne, Algebraic Geometry, Springer Graduate Texts #52, New York, 1977.

[Ig] J. -I. Igusa, Some observations on higher degree characters, Amer. J. Math., 99 (1977) 393–417.

[IR] K. Ireland and M. I. Rosen, Elements of Number Theory, Bogden & Quigley, Tarrytown, NY, 1972.

[J] J. -R. Joly, Equations et variétés algébriques sur un corps fini, Enseign. Math., 19 (1973) 1–117.

[O] G. Orzech and M. Orzech, Plane Algebraic Curves, Marcel Dekker, New York, 1981.

[R] A. Robert, Elliptic Curves, Springer Lecture Notes #326, New York, 1973.

[Sch] W. M. Schmidt, Equations over Finite Fields an Elementary Approach, Springer Lecture Notes #536, New York, 1976.

[Se] J. -P. Serre, A Course in Arithmetic, Springer Graduate Texts #7, New York, 1973.

[Si_1] W. Sierpiński, On some unsolved problems of arithmetics, Scripta Math., 25 (1960) 125–136.

[Si_2] ______, Remarques sur le travail de M. J. W. S. Cassels "On a diophantine equation," Acta Arith., 6 (1961) 469–471.

[S] C. Small, When is a power series rational?, preprint, 1979.

[T] J. T. Tate, The arithmetic of elliptic curves, Invent. Math., 23 (1974) 179–206.

NOTES

EDITED BY SHELDON AXLER, KENNETH R. REBMAN, AND J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057

A SPACE WHICH WILL NOT GO WHERE ITS PROPER SUBSPACES WANT IT TO GO

MURRAY BELL
Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

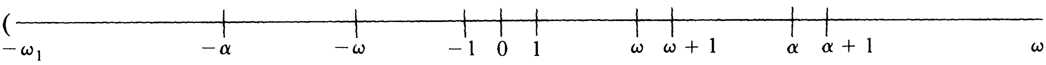
JOHN GINSBURG
Department of Mathematics, University of Winnipeg, Winnipeg, Manitoba, Canada R3B 2E9

In this note we describe an example of a pair of topological spaces X and Y such that every proper subspace of X can be embedded in Y, but X itself cannot be embedded in Y. A familiar example of spaces having these properties is a circle X and a line Y. In this context the two key properties of the circle are that it is homogeneous (for every pair of points p, q of the space, there is a homeomorphism of the space onto itself which maps p to q), and it is not homeomorphic to any proper subspace of itself. Here we will exhibit an *ordered* space X having the same two properties. Then take $Y = X - \{p\}$ for any point p in X: X cannot be embedded in Y since X is not homeomorphic to any of its proper subspaces, but every proper subspace of X can be embedded in Y since $X - \{q\}$ is homeomorphic to Y for every q in X.

We recall that every totally ordered set $(T, \leqslant)$ may be regarded as a topological space in a natural way; the open intervals in T are taken as a basis for the topology. Such an ordered space T is connected if and only if its underlying ordering is Dedekind complete and contains no consecutive elements, and T is compact if and only if its underlying ordering is complete. (These facts concerning ordered spaces may be found in 3-0 of [**1**].)

Let us now construct the space X, which is related to a well-known ordered space called the long line. X is in some sense doubly long. We begin by reviewing the construction of the long line.

Let ω_1 denote the first uncountable ordinal with its usual ordering. Thus the elements of ω_1 are the countable ordinals. For every $\alpha \in \omega_1$ the set $\{\xi \in \omega_1 : \xi < \alpha\}$ is countable, and if S is any countable subset of ω_1, there exists an element α in ω_1 such that $x < \alpha$ for all x in S. The long line L is formed by inserting copies of the Euclidean interval $(0, 1)$ between every pair of consecutive elements of ω_1. (Technically L is $\omega_1 \times [0, 1])$ ordered lexicographically: $(\alpha, s) \leqslant (\beta, t)$ if either $\alpha < \beta$ or if $\alpha = \beta$ and $s \leqslant t$.) This creates a connected ordered space L from the disconnected space ω_1. Let L^* denote the dual of L. That is, L^* is the set L with its ordering reversed. Let $M = L^* - \{0\}$, and let X be the ordinal sum of M and L. That is, X is $M \cup L$ ordered as follows: every element of M precedes every element of L and M and L retain their original orders. X can be pictured as follows, where L^* is thought of as $\{-x : x \in L\}$.



We claim that the ordered space X has the desired properties. First, X is Dedekind complete and contains no consecutive elements, hence is connected. To show that X is homogeneous, we first show that if x, y are in X and $x < y$, then the closed interval $[x, y]$ in X is homeomorphic to the Euclidean interval $[0, 1]$. This is seen as follows. We note that $[x, y]$ has a countable basis—we take a countable neighborhood base at each of the countably many ordinals and "dual ordinals" which lie in $[x, y]$, together with a countable neighborhood base at each rational number in the countably many copies of $(0, 1)$ between the consecutive ordinals or dual ordinals of $[x, y]$, together with a countable neighborhood base at x and y. These together form a countable basis for $[x, y]$. Since $[x, y]$ has a countable basis, it is a metrizable space. Also since $[x, y]$ is complete and has no consecutive elements, $[x, y]$ is compact and connected. Now any compact, connected metric space having exactly two noncut points is homeomorphic to $[0, 1]$. (By a noncut point of a space T is meant a point $p \in T$ such that $T - \{p\}$ is connected. A proof of the preceding characterization of $[0, 1]$ may be found, for example, in [**2**].) Since $[x, y]$ has exactly two noncut points, x and y, it follows that $[x, y]$ is homeomorphic to $[0, 1]$. Now we can show that X is homogeneous: Let x and y be any two points of X. Let α be an ordinal large enough that both x and y are inside the open interval $(-\alpha, \alpha)$. Since $[-\alpha, \alpha]$ is homeomorphic to $[0, 1]$, there is a homeomorphism of $[-\alpha, \alpha]$ onto itself which takes x to y. Such a homeomorphism must take end-points to end-points, and so it is clear that such a homeomorphism can be extended to a homeomorphism of X onto itself.

Finally, we establish the most interesting property of X, namely that X is not homeomorphic to any proper subset of itself. For, if Y is a subspace of X which is homeomorphic to X, then Y must be connected since X is. Now any connected subset of an ordered set must be an interval, so Y is an interval in X. Also Y is homogeneous, being homeomorphic to X, and so Y cannot have end-points. Thus Y is of one of the following forms:

$$(-\omega_1, \beta), \quad (\alpha, \beta), \quad (\alpha, \omega_1), \quad \text{or} \quad (-\omega_1, \omega_1).$$

It remains to show that X is not homeomorphic to any interval of the first three types. Now (α, β) is homeomorphic to the Euclidean interval $(0, 1)$, as seen above. But X has no countable dense subset since any countable subset of X is bounded in X. Therefore X is not homeomorphic to (α, β). We now show that X is not homeomorphic to (α, ω_1); the proof for $(-\omega_1, \beta)$ is similar. We will show that X *cannot* be expressed as $X = A \cup B$ where A and B are disjoint, nonempty, connected sets and where A has a countable dense subset. Since (α, ω_1) *can* be expressed in that way, as $(\alpha, \omega_1) = (\alpha, \beta) \cup [\beta, \omega_1)$ for any $\beta > \alpha$, it follows that the two spaces are not homeomorphic. To prove that X cannot be decomposed in this manner, assume $X = A \cup B$ where A and B are disjoint, nonempty, connected subsets of X. Then both A and B are intervals in X. Let $x \in A$ and let $y \in B$. If $x < y$, then no element of B can be less than x. (For, if $z \in B$ and $z < x$, then $z < x < y$, and it would follow that $x \in B$ since B is an interval.) Therefore $(-\omega_1, x) \subseteq A$. Similarly if $y < x$, then $(x, \omega_1) \subseteq A$. A similar argument shows that either $(-\omega_1, y) \subseteq B$ or $(y, \omega_1) \subseteq B$. Therefore both A and B are unbounded intervals and so neither has

a countable dense subset. (This follows from the fact that every countable subset is bounded.) Thus X admits no decomposition of the form $X = A \cup B$ where A and B are disjoint, nonempty, connected sets, such that A contains a countable dense subset. This completes the proof.

References

1. L. Gillman and M. Jerison, Rings of Continuous Functions, Van Nostrand, Princeton, N.J., 1960.
2. S. Willard, General Topology, Addison-Wesley, Reading, Mass., 1970.

A GEOMETRIC APPROACH TO INVARIANT SUBSPACES OF ORTHOGONAL MATRICES

PETER BUSER

Department of Mathematics, State University of New York, Stony Brook, NY 11794

An $n \times n$ orthogonal matrix A with real entries can be put into a normal form consisting of 2×2 and 1×1 blocks. This is equivalent to showing that R^n splits into an orthogonal sum of 1- or 2-dimensional subspaces which are invariant under the action of A. The usual argument proceeds by induction from one subspace and uses facts about factoring polynomials. Here is a purely geometrical argument which yields one such subspace for the induction process.

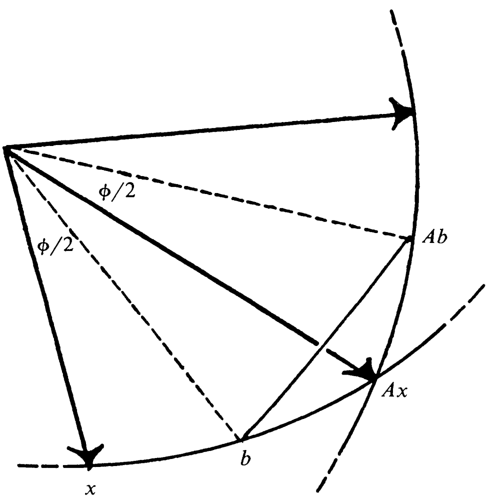


FIG. 1.

If the orthogonal matrix A has real eigenvalues, we are done. If not, for some $x \in R^n$, x and Ax are not parallel; these vectors determine a plane. Since the angle between x and Ax may be thought of as a distance on the compact surface of the unit sphere, there is an $x \in R^n$ for which the angle ϕ between x and Ax is a minimum. Let b be the bisector of the angle between x and Ax (any other vector in the same plane would serve). Being orthogonal A preserves angles, hence the triangle inequality implies

$$\phi \leqslant \sphericalangle(b, Ab) \leqslant \sphericalangle(b, Ax) + \sphericalangle(Ax, Ab) = \phi/2 + \phi/2.$$

Clearly the inequalities must be equalities which forces Ab to lie in the same plane as b, x, and AX. Hence x, b span a two-dimensional invariant subspace.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

EXISTENCE STATEMENTS AND CONSTRUCTIONS IN MATHEMATICS AND SOME CONSEQUENCES TO MATHEMATICS TEACHING

SHLOMO VINNER
Israel Science Teaching Center, Hebrew University, Jerusalem, Israel

DAVID TALL
Mathematics Education Research Center, University of Warwick, Coventry, England

There are many axioms, theorems and constructions in mathematics the role of which is to guarantee the existence of certain mathematical objects. (For instance, an axiom: there exists an infinite set; a theorem: every angle has a bisector; a construction: the construction of the rational numbers as equivalence classes of pairs of integers.) These will be discussed from philosophical, mathematical and psychological points of view. Some consequences to the instruction of topics involving existence statements and constructions will be drawn.

1. The Notion of Mathematical Objects and the Realist Approach. The mathematical community speaks about mathematical objects. Numbers, functions and geometrical shapes, for instance, are among many other things which are considered as mathematical objects. The expression "mathematical objects" perhaps does not occur very often in the language of mathematicians but you can hardly find a mathematician who has not used it at least once or twice. The word "object" has the connotation of something which exists outside our mind. In fact, there are people who believe that mathematical objects have an existence which is independent of our mind. This belief (which has several versions with different degrees of sophistication) can be considered as the basis for Realism (sometimes also call Platonism) in mathematics(1). Hence, a mathematician who speaks about mathematical objects acts as a realist. By this we do not claim that he really believes that mathematical objects actually exist outside our mind (as we do not claim that somebody who utters, "Oh, my God," believes that God exists). One can use quite sophisticated arguments to show that the contrary is true. Our only claim is that all mathematicians at least once or twice act *as if* they were realists (exactly as every human being acts, at least once or twice in his life, as if he believes in God).

The Realist approach to mathematical objects is, in a way, the simplest approach. If you are willing to accept without questioning nonconcrete objects which exist outside your mind, then you will not face further difficulties. Our impression is that many mathematicians implicitly accept it, but the moment they have to admit it in public they deny it. This is, probably, because it is hard to explain how something which is not concrete exists outside our mind. However, the basis for other approaches to mathematical objects is the claim that the expression "mathematical objects" is a "*façon de parler*" and it is not at all an evidence that we believe there are unconcrete things which exist outside our mind (exactly as the expression, "Oh, my God," is a "*façon de parler*" for the atheist).

2. The Psychological Need for Mathematical Realities, A Psychological Point of View. There is no point in making statements unless you believe that these statements relate to a certain reality.

(1) The reader who is unfamiliar with these notions is advised to read [6] which has an excellent survey on the main philosophical approaches to mathematics and also a further list of references on this topic.

The moment a statement is made the mind becomes busy with constructing, reconstructing, recalling, evoking or imagining realities. We consider this a result of a very basic need of our cognition, *the need for realities to which statements refer.* Thus a fundamental question in mathematics education, although an implicit one in many cases, is the question of mathematical realities.

Any scientist, not being in a too philosophical mood, is convinced that the scientific statements he makes relate to the real world(2). Mathematics, on the other hand, does not relate to the real world. However, the *need for realities creates realities in the mind*, the realities of mathematical objects. If there is no real world to which the mathematical statements relate, then an imaginary world is constructed according to explicit or implicit guidelines of the mathematical context. For instance, mathematical objects like points, lines and planes do not exist in the real world. But some real world figures are used as a starting point for the *imagination constructs.* These constructs will finally form the mathematical objects which will constitute the mathematical reality, in this case—the geometrical space. In order to construct the geometrical space, cognition should act in a certain way. It should, perhaps, start with a real point, then *imagining* it becoming thinner and thinner until finally, somehow, it will have no dimensions at all. It should start with an actual line segment, perhaps drawn on a sheet of paper, then extending it (in the mind) to both sides, finally *imagining* that it has no ends.

Let us call mind activities of this type by the name "*imagination acts.*" Imagination acts are idiosyncratic experiences. Fortunately enough, many people share similar idiosyncratic experiences which become the basis for mathematical communication. Hence, the fact that mathematicians can communicate is not due to the fact that there are mathematical realities which exist outside their mind. It is due to the fact that the imagination acts of many different people have a similar structure. But this fact is a psychological fact. Using it as a starting point, we can suggest that mathematical realities are constructs in the mathematicians' mind. Thus, we are trying to solve the problem of mathematical realities within the framework of psychology. As mathematics educators we cannot ignore psychology, but this seems quite unsatisfactory for many mathematicians who feel that the problem should be solved within the framework of mathematics. Their best option is the Formalist approach.

3. The Formalist Approach. The essence of formalism in mathematics is the elimination of all "mathematical objects." Mathematics is considered merely from the linguistic point of view as a formal system. It is a field where theorems are inferred from axioms by means of inference rules and there is no reality involved(3). The only problem in this approach is to prove the consistency of mathematics, namely, to prove that it is impossible to derive a contradiction from the axioms of mathematics by means of the inference rules. This is in fact one of the main ideas behind the Formalist Program, known as Hilbert Program. It was a call to mathematicians to prove the consistency of mathematics within the restricted framework of mathematics and thus to eliminate the need for mathematical realities. This does not imply that mathematicians are not allowed to think in terms of mathematical realities, especially if it helps them to do mathematics. It only implies that mathematics itself as a collection of mathematical statements does not depend on mathematical realities which are quite problematic, as we claimed before. Unfortunately, it is impossible to carry out the Hilbert Program as was proved by Gödel in 1931 (see [**2**]). Namely, it is impossible to prove the consistency of mathematics within its restricted framework.

4. The Meaning of Mathematical Existence from the Mathematical Point of View. According to the Formalist approach to claim that certain mathematical objects exist means to claim that there are axioms or theorems that (explicitly or implicitly) assert the existence of these mathemati-

(2) There are some problems even here, but since our main concern in this paper is pedagogy and not philosophy, we will avoid philosophical sophistication. The interested reader can consult a philosophy of science text, such as [**3**].

(3) For more details see [**6**] again.

cal objects. For instance, to claim that for any sets X and Y there exists a set Z such that for every x: $x \in Z$ if and only if $x \in X$ or $x \in Y$ means that this statement is an axiom or a theorem of certain mathematical theory (in this case it is an axiom in one of the well-known axiom systems of set theory. See, for instance, [1]). Not always are the axioms or theorems that assert existence so explicit or direct. For instance, to claim, within the restricted framework of mathematics, that the real numbers exist means that there is a construction of some mathematical objects, and there are theorems asserting that these objects have certain properties (in this case the well-known properties of the real field). As a matter of fact, mathematics textbooks do not contain theorems like "the real numbers exist." Nevertheless, this is implicitly implied by the detailed and sophisticated constructions. They all come to establish the existence of certain mathematical realities.

However, any claim that mathematical existence implies more than mentioned above (particularly, that it has something to do with mathematical realities) is not within the restricted framework of mathematics. It implies a philosophical standpoint such as the Realist view. Thus, if you, in spite of all, insist on staying within the restricted framework of mathematics, then (whether you are a Formalist or not) you must adopt the Formalist approach to existence in mathematics.

5. Relative versus Absolute Existence. At first sight there is a clear difference between the two following existence statements:

1. For any two given sets there exists a set which is their union.
2. The real numbers exist.

The first statement claims the existence of certain objects in a world the existence of which is unquestionable in this context. Hence, we will call such an existence a relative existence. Here, the existence of a whole reality is presumed before the claim of the particular existence is made.

On the other hand the claim that the real numbers exist is a claim about a whole system. It looks much stronger than the claim about the union set and at first sight it does not seem to be relative in the sense stipulated above. An existence which is not claimed within the framework of other systems will be called absolute.

However, considering the claim about the real numbers more closely, you recall immediately that the real numbers are constructed by means of the rational numbers (whether Dedekind cuts, Cantor sequences or Cauchy sequences have been used). Hence, in order to construct the real numbers, one needs to assume the existence of the rational numbers. Of course, the rational numbers can also be constructed, but in order to do that one has to assume the existence of the natural numbers. At this point one should probably admit an unquestionable existence of the system of the natural numbers[4]. This is really an absolute existence, but admitting it is either a philosophical or psychological act, not a mathematical one. This situation is beautifully described by the famous statement of Kronecker: "God created the natural numbers. All the rest was made by Man." According to Kronecker the existence of the natural numbers cannot be guaranteed by mathematicians. For this purpose, at least, God is needed. In other words, mathematicians cannot prove *absolute existence*. It is only the relative existence that they prove and discuss. No absolute existence of mathematical realities can be proved within the restricted framework of mathematics.

6. The Psychological Need for Mathematical Realities and the Offer of Mathematics. If our description above is correct, then the psychological need for mathematical realities cannot be satisfied within the restricted framework of mathematics. There are mainly two reasons for this. First, mathematics can offer only a relative existence. But what one needs is an absolute existence. (One needs an unquestionable reality with which mathematics is associated. Mathematics estab-

[4] As a matter of fact, one can go further and construct the natural numbers within Set Theory as it was done in [5], but then the existence of the world of sets should be admitted and this will lead to the same conclusion that will immediately be drawn.

lishes, for instance, irrational numbers *on* rational numbers, but in any reality we see one kind of objects *at the side* of another kind.)

Second, the principles according to which imagination acts (§2) take place are different from the principles by means of which mathematics is developed. (Cognition needs special training in order to act according to some principles of mathematics and this training is hard and long.) One of these principles can be called the principle of familiarity. Cognition seeks familiarity. It expects new things to be similar to things which it already knows. It tries to interpret new concepts in terms of familiar ones. If numbers, for instance, are associated with quantities, proportions or length, then one expects any new treatment of the concept of number to be related somehow to these notions. Of course, this is not the case. In order to get a new point of view, cognition should accommodate. A major accommodation might take months or even years. (According to Piaget, [4], it takes about two years until a child develops from one intellectual stage to the next one.) During this time one should face, live and cope with many experiences which finally make this accommodation happen.

7. The Right Time for Existence Statements and Constructions. From what we have said it follows that existence statements and constructions can help the undergraduate student very little. If we are concerned with existence statements which seem trivial (such as every segment has a midpoint, an area function exists, every finite set of numbers has a minimal element), then there is no need to convince the student that they are true. He is convinced the moment these statements are made. Proving them can be meaningful only if the student realizes that proof and conviction in mathematics are two different things, that mathematical truths should be established according to certain formal rules while intuition, self-evidence and beliefs can play only an external role. This requires a major accommodation.

On the other hand, if we are concerned with sophisticated existence constructions, then the requirement of familiarity (§6) is not fulfilled. The student previously formed certain mathematical realities in his mind and suddenly he has to think, for instance, that numbers are equivalence classes of integers or Dedekind cuts or Cantor sequences. Again, this requires a major accommodation. The only candidates for such an accommodation can be mathematics major students at a senior stage of their studies (first year of graduate studies is perhaps the best). At this stage they are psychologically and intellectually matured to question mathematical realities and to understand the point of the Formalist approach to Mathematics. Undergraduate students, especially those who are not mathematics majors should be taught in such a way that major accommodation will not be needed. For these students intellectual conflicts are not productive. They are embarrassing and upsetting[(5)].

It is well known that some mathematics departments use existence statements and constructions (whether they are part of various courses or they constitute a whole course like Number Systems) as a means of selection. Knowing that existence statements and constructions are quite hard, they introduce them to the students at an early stage, expecting the weak students to fail. Of course, every department has the right to select its students. One can argue whether this means is the appropriate one. But from the educational point of view, existence statements and constructions, when taught at an inappropriate stage, are a total waste of time and energy.

References

1. A. A. Fraenkel, Y. Bar-Hillel, and A. Levy, Foundations of Set Theory, North-Holland, Amsterdam, 1966.

2. K. Gödel, On Formally Undecidable Propositions of Principia Mathematica and Related Systems, in J. V. Jeijenoort, From Frege to Gödel, Harvard Univ. Press, Cambridge, 1967.

3. C. H. Hempel, Philosophy of Natural Sciences, Prentice-Hall, Toronto, 1966.

4. J. Piaget and B. Inhelder, The Growth of Logical Thinking, Routeledge and Paul Kegan, London, 1958.

(5) In [7] another suggestion is made to avoid the need of major accommodation in a similar situation in the learning of mathematics.

5. B. Russell and A. N. Whitehead, Principia Mathematica, Cambridge Univ. Press, Cambridge, 1910.
6. E. Snapper, The three crises in mathematics: logicism, intuitionism and formalism, Math. Mag., vol. 52, no. 4 (1979) 207–216.
7. S. Vinner, Implicit axioms, ω-rule and the axiom of induction in high school mathematics, this MONTHLY, August-September (1976) 561–566.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better, If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by April 30, 1983. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2974. *Proposed by Jordi Dou, Barcelona, Spain.*

Let AMB (oriented clockwise) and CMD (counterclockwise) be similar triangles. Prove that triangles ACX (clockwise) and YDB (counterclockwise), both similar to the first triangles, have the same circumcenter.

E 2975. *Proposed by F. Burton Jones, University of California, Riverside.*

An arc is a one-to-one reversibly continuous image of $[0,1]$ of the real numbers. Does there exist an arc in the plane whose 2-dimensional Lebesgue measure is positive?

E 2976. *Proposed by Lee Whitt, D. H. Wagner Associates, Hampton, Virginia.*

Let N_0 be a given nonnegative integer and p be a real number such that $0 < p < 2$. Give a closed form expression for the summation

$$\sum_{L=N_0}^{\infty} \binom{L}{N_0} p^{N_0}(1-p)^{L-N_0}$$

where we define $0^0 = 1$ for $p = 1$ and $L = N_0$. Note that the solution is independent of N_0.

E 2977. *Proposed by Stan Wagon, Smith College.*

Martin Gardner once mentioned the following problem in his *Scientific American* column:

Partition the positive integers into two sets, A, B, such that neither $A + A$ nor $B + B$ contains a prime. ($X + X$ denotes $\{x + y\colon x, y \in X, x \neq y\}$.) Show that there is a *unique* solution to this problem. (See advanced proposal 6413, page 788, in this issue.)

E 2978. *Proposed by U. Abel, DKFZ, Heidelberg, Germany.*

A mapping g is called cyclic if $P(g) = \min\{k \mid g^k = \text{identity}\}$ exists ($g^k = g \circ \cdots \circ g$, k times).

Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a cyclic linear mapping, $R^n = V \oplus W, V, W$ being subspaces of $\mathbb{R}^n$, and $f(V) = V$.

Define $\tilde{f} = \mathrm{Pr}_W \circ f|_W\colon W \to W$ (Pr_W denoting the projection on W along V). Then $\tilde{f}$ is cyclic, and $P(f) = \mathrm{lcm}(P(\tilde{f}), P(f|_V))$.

E 2979. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let $\{p_k\}$ and $\{q_k\}$ $(k = 0, 1, 2, \dots)$ be two sequences of real, monic, orthogonal polynomials, p_k and q_k having degree k. Show that for any nonnegative integers m and n, the polynomial $p_m q_{n+1} + p_{m+1} q_n$ has only real simple zeros.

SOLUTIONS OF ELEMENTARY PROBLEMS

Primitive Roots Modulo p

E 2488 [1974, 776]. *Proposed by Richard Stanley, University of California, Berkeley.*

Let p be an odd prime. It has been conjectured that there exists a natural number $k \leqslant p - 1$ which is a primitive root modulo p and which is relatively prime to $p - 1$. Prove this conjecture in the special case that $p \equiv 1 \pmod 4$ and $3\phi(p - 1) > p - 1$, where ϕ denotes Euler's totient function.

The solution to this problem appeared in this MONTHLY [vol. 83, 1976, pp. 720–723].

The Equation $[an] + [bn] = [cn] + [dn]$

E 2752 [1979, 56] *Proposed by Clark Kimberling, University of Evansville.*

Let a, b, c, d be real numbers such that, for all $n > n_0$, the relation (**) $[an] + [bn] = [cn] + [dn]$ holds. (i) Prove that $a + b = c + d$. (ii) Prove that $[a] + [b] = [c] + [d]$. (iii)* Find all solutions of (**).

Solution to (i) by Robert Breusch, Amherst, Mass. If x is real, the relation $x - 1 < [x] \leqslant x$ holds. Divide (**) by n and let $n \to \infty$, to reach the conclusion $a + b = c + d$.

Solution to (ii) by T. Jager, Calvin College. Write

$$a = A + \alpha, \qquad b = B + \beta, \qquad c = C + \gamma, \qquad d = D + \delta, \qquad 0 \leqslant \alpha, \beta, \gamma, \delta < 1;$$

$$A, B, C, D \text{ integers } (\alpha, \beta, \gamma, \delta \text{ are the fractional parts of } a, b, c, d).$$

If $A + B = C + D$, there is nothing to prove.

The other possibilities are

$$(^{***})\ A + B + 1 = C + D, \qquad \alpha + \beta = \gamma + \delta + 1;$$

$$(^{***\prime})\ A + B = C + D + 1, \qquad \alpha + \beta + 1 = \gamma + \delta.$$

If (***) holds, then

$$[na] + [nb] - [nc] - [nd] = n + [n\alpha] + [n\beta] - [n\gamma] - [n\delta].$$

But

$$[n\alpha] + [n\beta] > [n(\alpha + \beta)] - 2 = [n(\gamma + \delta + 1)] - 2$$
$$= [n(\gamma + \delta)] + n - 2 \geqslant [n\gamma] + [n\delta] + n - 2,$$

so that $[na] + [nb] - [nc] - [nd] > 2n - 2$. If $n > 0$, this contradicts (**). Thus the relation $A + B = C + D$ is established.

Before going on to (iii), we note that any integer N can be added to a, if the same integer is subtracted from b, without changing the value of $[na] + [nb]$. From (i) and (ii), it follows that the only case to discuss is the case $0 \leqslant a, b, c, d < 1$. The only subcase of interest is the case ($*^{\dagger}$) $0 \leqslant a < c \leqslant d < b < 1$. (Otherwise, (c, d) is just a permutation of (a, b).)

Comments by Floyd Barger, Youngstown, Ohio, Eli L. Isaacson, New York, NY, and Herb Shank, Waterloo, Canada (independently). First, if a, b, c, d are all irrational, and if $a + b = c + d = 1$, then (**) holds. *Proof.* Let ny be not an integer. Then $[n - ny] = n - 1 - [ny]$. Thus $[na] + [nb] = [na] + [n(1 - a)] = n - 1$. □

Second, if a, b, c, d are rational in lowest terms with the same denominator, that is, if $a = A/M$, $b = B/M$, $c = C/M$, $d = D/M$, with $(A, M) = (B, M) = (C, M) = (D, M) = 1$, and if furthermore $a + b = c + d = 1$, then (**) holds. *Proof.* Use the above comment, and the additional observation that if nx is an integer, then $[nx] + [n(1 - x)] = n$. □

Thus (iii) is partially answered: some solutions have been found. It remains to show that there are no other solutions. The editors can handle the irrational cases (1°) – (4°) below, using (a generalization of) a theorem of Kronecker.

(1°) One of the four sets $\{a, b, c, 1\}$, $\{a, b, d, 1\}$, $\{a, c, d, 1\}$, $\{b, c, d, 1\}$ is linearly independent over the rationals.

(2°) The numbers a, b are both irrational, but the numbers c, d are both rational.

(3°) The numbers a, c are irrational, b, d are rational, $0 \leqslant a, b, c, d < 1$, $b = R/T$, $d = S/T$, $b \neq d$, $0 < a + b = c + d < 1$.

(4°) Certain subcases under the case a, b, c irrational, d rational.

The case (5°) all four of the sets $\{a, b, c, 1\}, \ldots, \{b, c, d, 1\}$ are linearly dependent over the rationals, a, b, c, d all irrational, $0 < a + b = c + d < 1$ remains open. Perhaps the method L. L. Foster uses in proving the result below will be useful in this case.

(6°) The rational case: a, b, c, d all rational, but one or more of the conditions (above in "Second") fails. That is, either $0 < a + b = c + d < 1$, or else $a + b = c + d = 1$, but a, b, c, d (in lowest terms) do not all have the same denominator.

Solution by Lorraine L. Foster, California State University at Northridge. In the last case, e.g., $a = 1/3$, $b = 2/3$, $c = 2/5$, $d = 3/5$, it is obvious that (**) fails for infinitely many values of n. We proceed to the hard case.

The fractional part of x, $x - [x]$, is denoted $\{x\}$. The result proved concerning rational a, b, c, d, $a = A/M$, $b = B/M$, $c = C/M$, $d = D/M$, is the following.

THEOREM. *Let A, B, C, D, M be positive integers such that the conditions* (i) $A \neq C$, $A \neq D$, (ii) $\gcd(A, B, C, D, M) = 1$, (iii) $0 < A, B, C, D < M$, (iv) $A + B = C + D \neq M$ *all hold. Then there exist infinitely many positive integers $n > 0$ such that the inequality $[nA/M] + [nB/M] \neq [nC/M] + [nD/M]$ is valid.*

This theorem is proved by establishing four lemmas.

LEMMA 1. *Suppose the positive integers A^*, B^*, C^*, D^*, M^* satisfy* (i*) $A^* \neq C^*$, $A^* \neq D^*$, (ii*) $\gcd(A^*, B^*, C^*, D^*, M^*) = 1$, (iii*) $0 < A^*, B^*, C^*, D^* < M^*$, (iv*) $A^* + B^* = C^* + D^* \neq M^*$. *Further, suppose that, for all $n > n_0 > 0$, the relation*

$$[nA^*/M^*] + [nB^*/M^*] = [nC^*/M^*] + [nD^*/M^*] \qquad (1)$$

[Continued on p. 779.]

holds. Then there exist positive integers A, B, C, D, M *that satisfy* (i) $A \neq C, A \neq D$, (iii) $0 < A, B, C, D < M$, (iv) $A + B = C + D \neq M$, (v) $(A, M) = (C, M) = 1$, (vi) $(B, M) = (D, M) = d \geqslant 1$, *and for all* $n > n_1$, *the relation*

$$[nA/M] + [nB/M] = [nC/M] + [nD/M] \tag{2}$$

holds.

Proof. The relation (1) is given. Set $(A^*, M^*) = d_1 \geqslant 1, (B^*, M^*) = d_2 \geqslant 1$. Let $q > 1$ be such that $q \mid A^*$, $q \mid B^*$, $q \mid M^*$. Apply (1) with

$$n = kM^*/q;\ k = k_0, k_0 + 1, \ldots: \{kA^*/q\} + \{kB^*/q\} = 0, k = k_0, k_0 + 1, \ldots.$$

If k is chosen relatively prime to q, this last relation implies that $q \mid C^*$, $q \mid D^*$, which contradicts (ii*). Therefore $q > 1$ does not exist, $(d_1, d_2) = 1$, and we can write $M^* = D_1 D_2 K$, where $(K, A^*B^*) = 1$, and where

$$D_1 = \Pi_{p_i \mid d_1} p_i^{\alpha_i}, \qquad D_2 = \Pi_{q_i \mid d_2} q_i^{\beta_i},$$

where p_i, q_i are some primes, where the exponents α_i, β_i are positive, and where $(D_1, D_2) = 1$.

To continue the reduction, suppose $d_2 > 1$ and choose the integer $X > n_0$ so that $XA^* \equiv 1 \bmod d_2$. Then the relation $\{XA^*/d_2\} + \{XB^*/d_2\} = 1/d_2$ holds automatically. Now $XC^* + XD^* = XA^* + XB^*$; w.l.o.g. $d_2 \mid D^*$. Then clearly

$$(B^*, M^*) \mid (D^*, M^*). \tag{3_1}$$

Also $XA^* \equiv 1 \equiv XC^* \bmod d_2$ for some $X > n_0$, so that

$$A^* \equiv C^* \bmod d_2. \tag{3_2}$$

Clearly $(3_1, 3_2)$ holds even if $d_2 = 1$.

Similarly, either $d_1 \mid C^*$ and $(d_1, D^*) = 1$, or else $(\dagger) d_1 \mid D^*$ and $(\dagger\dagger)(d_1, C^*) = 1$. Suppose the last $(\dagger, \dagger\dagger)$ hold. If further $d_1, d_2 > 1$, then we can choose $Y > n_0 d_1 d_2$ such that $YC^* \equiv 1 \bmod d_1 d_2$. Hence, putting $n = YM^*/(d_1 d_2)$ in (1), it would follow that

$$\{YC^*/(d_1 d_2)\} + \{YD^*/(d_1 d_2)\} = 1/(d_1 d_2) < \{YA^*/(d_1 d_2)\} + \{YB^*/(d_1 d_2)\},$$

which cannot be the case. Thus $(\dagger, \dagger\dagger)$ is contradicted, and we have shown that $d_1 \mid C^*, (d_1, D^*) = 1$. This is also true if $d_1 = 1$. Thus

$$(A^*, M^*) \mid (C^*, M^*), B^* \equiv D^* \bmod d_1. \tag{4}$$

The final case $d_2 = 1, d_1 > 1$ does not present any new difficulty; the roles of d_1, d_2 can be reversed in the above argument. Thus, by renaming the symbols if necessary, we have from (3), (4):

$$(A^*, M^*) = (C^*, M^*), \qquad (B^*, M^*) = (D^*, M^*), \qquad A^* \equiv C^* \bmod d_2,$$
$$B^* \equiv D^* \bmod d_1, d_1, d_2 \geqslant 1.$$

If $d_1 \doteq 1$, the truth of the lemma follows with

$$A = A^*, \qquad B = B^*, \qquad C = C^*, \qquad D = D^*, \qquad M = M^*, \qquad d = d_2.$$

If $d_1 > 1, d_2 = 1$, the truth of the lemma follows with

$$B = A^*, \qquad A = B^*, \qquad C = D^*, \qquad D = C^*, \qquad M = M^*, \qquad d = d_1.$$

Finally, suppose $d_1 > 1, d_2 > 1$. In this case it is automatic that $A^* \not\equiv D^* (\bmod D_2 K)$, $B^* \not\equiv C^* (\bmod D_1 K)$. We observe that either $A^* \not\equiv C^* \bmod D_1 K$ or else $A^* \not\equiv C^* \bmod D_2 K$. For suppose the contrary $A^* \equiv C^* \bmod D_1 D_2 K$. Then $B^* \equiv D^* \bmod D_1 D_2 K$. This carries with it $A^* \equiv C^* \bmod M^*$, so that $A^* \equiv C^*$, contradicting one of the hypotheses.

With $A^* \not\equiv C^* \bmod D_1 K$, note $B^* \not\equiv D^* \bmod D_1 K$. Define A', B', C', D' as the least nonnega-

tive residues of A^*, B^*, C^*, D^* mod D_1K. Then (see below) the lemma is true with $A = B'$, $B = A'$, $C = D'$, $D = C'$, $M = D_1K$.

With $A^* \not\equiv C^*$ mod D_2K, define A', B', C', D' similarly, but using D_2K in place of D_1K, thus: $A = A'$, $B = B'$, $C = C'$, $D = D'$, $M = D_2K$. In either case, we have $A \neq 0, C \neq 0, A \neq C, A \neq D$.

To complete the proof of the lemma in the case of the preceding paragraph, we show that $B \neq 0$. Suppose, contrariwise, that $B = 0$ in that paragraph. Choose $X' > n_0 D_2 K$ such that $X'A \equiv 1 \bmod D_2K$. Then

$$\{X'A/(D_2K)\} + \{X'B/(D_2K)\} = 1/(D_2K),$$

so that, mod D_2K,

$$D \equiv B \equiv 0, \qquad X'A \equiv X'C, \qquad A \equiv C,$$

which contradicts one of the hypotheses, since $D_2 > 1$. Thus $B \neq 0$, and by symmetry $D \neq 0$. Also, $A + B \neq D_1K$ because $d_1 \nmid B$. This establishes (2).

In the other case $A^* \not\equiv C^*$ mod D_1K, the argument is similar; we find $BD \neq 0$, $A + B \neq D_2K$ in this case. □

Lemma 2. *Suppose A, B, C, D satisfy* (i) $A \neq C$, $A \neq D$, (iii) $0 < A, B, C, D < M$, (iv) $A + B = C + D$, (vi) $(A, M) = (C, M) = 1$, $(B, M) = (D, M) = d \geqslant 1$. *Let $X > Mn_0$ be such that $XA \equiv 1 \bmod M$; define B', C', D' as the respective least nonnegative (in fact positive) residues of XB, XC, XD mod M. Then the relation*

$$[nA/M] + [nB/M] = [nC/M] + [nD/M] \tag{5}$$

holds for all $n > n_0$ if and only if the relation

$$[w/M] + [wB'/M] = [wC'/M] + [wD'/M] \tag{6}$$

holds for all sufficiently large $w > w_1$.

Proof. Suppose that (5) holds. Take $n = X$; then from (iv), $1/M + B'/M = C'/M + D'/M$. Clearly $C' \neq 1$, $D' \neq 1$. Also, $1 + B' < M$ (since $A + B < 2M$; and $A + B \neq M$ by (vi), so that $X(A + B) \neq 0 \bmod M$).

Now take w positive. Then $\{wXA\} + \{wXB\} = \{wXC\} + \{wXD\}$. But

$$wXA \equiv w, \qquad wXB \equiv wB', \qquad wXC \equiv wC', \qquad wXD \equiv wD' \bmod M.$$

Thus for large w, relation (6) holds.

Conversely, suppose that (6) holds for all large $w > w_1$. For $w = LM$, we have $1 + B' = C' + D'$, so that automatically

$$w/M + wB'/M = wC'/M + wD'/M.$$

Further (with $w = nA$) we have

$$\{nA/M\} + \{nAB'/M\} = \{nAC'/M\} + \{nAD'/M\}$$

for integral n. But mod M, we note

$$nB \equiv nAXB \equiv nAB', \qquad nC \equiv nAC', \qquad nD \equiv nAD',$$

so that (since $nA + nB = nC + nD$) we have (5). □

Lemma 3. *Let s, t, u, M be positive integers and suppose that* (i) $u \neq s$, $u \neq t$, (iii) $s, t, u < M$, (iv) $1 + u = s + t$, (vii) $2 \leqslant t \leqslant s < M/2$, (viii) $2u \leqslant M$. *Suppose further that one of the three relations* (vi$_1$), (vi$_2$), (vi$_3$) *holds*:

$$(\mathrm{vi}_1) \quad (s, M) = (u, M) = d \geqslant 1, \quad (t, M) = 1;$$

$$(\mathrm{vi}_2) \quad (t, M) = (u, M) = d \geqslant 1, \quad (s, M) = 1;$$

$$(\mathrm{vi}_3) \quad (s, M) = (t, M) = d > 1, \quad (u, M) = 1.$$

Then there exists an integer $n, 0 < n < M/d$ (and hence there exist infinitely many integers n) such that the inequality

$$[n/M] + [nu/M] \neq [ns/M] + [nt/M] \tag{7}$$

holds.

Proof. Suppose first that (vi_3) holds, and that moreover $t = d$, $s = vd$, $v \geqslant 1$. Take $n = M/D - 1 \geqslant 1$. Then on the one hand $[n/M] \leqslant [nt/M] = [(M-d)/M] = 0$, and also $[ns/M] = [(Mv - s)/M] < v$. On the other hand, since $(d-1)/d - u/M \geqslant 1/2 - 1/2 = 0$, we have $[nu/M] = [(n(v+1)d - 1))/M] = [v + (d-1)/d - u/M] \geqslant v$. In this case, (7) is established.

Next if (vi_3) holds and $s > d$, $t > d$ (the only remaining possibility under (vi_3)), we define the sequences q_k, r_k $(k = 1, 2, \ldots)$ by means of the relations $kM = sq_k + r_k, 0 \leqslant r_k < s$. Let L denote the minimal subscript such that $(t-1)q_L \geqslant r_L$ (L surely exists). We define N by (the division algorithm) $s = tN + R, 0 \leqslant R \leqslant t - 1$. From the relation $(t-2)(N-1) \geqslant 0$, it follows that

$$tN + 1 \geqslant 2N + t - 1 \geqslant 2N + R;$$

using $s = tN + R$ again, we have $2tN \geqslant 2N + s - 1$. Thus $2N \geqslant (s-1)/(t-1)$, so that

$$[s/t]M = NM \geqslant 2Nu \geqslant u(s-1)/(t-1).$$

It is easy to show that the relation $(t-1)q_N < r_N$ cannot hold. For if it did, then it would follow that $r_N/(t-1) < q_N$, so that

$$NM = sq_N + r_N < (s + t - 1)r_N/(t-1) \leqslant u(s-1)/(t-1),$$

which contradicts the above. We may thus assume $(t-1)q_N \geqslant r_N$, so that

$$L \leqslant N \leqslant s/t, \; Lt/s \leqslant 1.$$

We can now settle case (vi_3), and fortuitously case (vi_1) as well. We note $d \mid s$, and set $s/d = x$, $N' = s/d - 1$. For (vi_3), since $s > d$, we note

$$x \geqslant 2, \; N' \geqslant 1, \text{ and } P = s + d + M/x \leqslant s + t - d + M/x \leqslant u + M/x \leqslant M.$$

For (vi_1), we have

$$t \equiv 1 \bmod d, \qquad t \geqslant d + 1, \qquad x \geqslant 2, \qquad N' \geqslant 1,$$

and again

$$P \leqslant s + t - 1 + M/x = u + M/x \leqslant M.$$

In (vi_3), (vi_1), we have $(s + M)(s + d)/s \leqslant 2M$. We next note that

$$N'M = (s/d - 1)M = s(M/d - y) + (ys - M),$$

where y is some integer such that

$$0 \leqslant ys - M < s, \quad \text{i.e.,} \quad y < (s + M)/s.$$

(Thus $r_{N'} = ys - M$.) But then, $y(s + d) < 2M$, and so $M - dy > ys - M$. For (vi_3), (vi_1), we have $(t-1)/d \geqslant 1$, so that (using $r_{N'} = ys - M$ twice)

$$(t-1)q_{N'} = ((t-1)/d)(M - dy) > ys - M = r_{N'}.$$

Thus, from the definition of L, $L < s/d$.

We now turn to (vi_2) and show that $L < s/d$ in this case also. We set $N'' = (s-1)/d$; this is an integer since $s \equiv 1 \bmod d$; obviously $N'' \geqslant 1$. Either $M \geqslant sd$ or $M < sd$. Suppose $M \geqslant sd$. Then $((s-1)/d)(M/s) \geqslant s - 1$, so that $[((s-1)/d)M/s] \geqslant s - 1$, and thus $q_{N''} \geqslant s - 1 \geqslant r_{N''}$. On the other hand, if $M < sd$, then $d \geqslant 2$, and

$$((s-1)/d)M = s((M-d)/d) + (sd - M)/d.$$

Since $0 < (sd - M)/d < s$, it follows that

$$q_{N''} = (M-d)/d, \qquad r_{N''} = (sd - M)/d.$$

From $d \mid t$, we see that

$$(t/d)M > u, tM > d(t+s-1), (t-1)(M-d) > sd - M,$$

and thus $(t-1)q_{N''} > r_{N''}$. Hence the stronger assertion $L \leqslant (s-1)/d$ holds. We have shown that $L < s/d$ in any one of cases (vi_1), (vi_2), (vi_3).

The next step in the argument is to establish the inequalities $q_L < M/d$, $r_L > 0$. Suppose, on the contrary, that $r_L = 0$, $LM = q_L s$. In cases (vi_1), (vi_3), this gives the immediate contradiction $(s/d) \mid L$. In case (vi_2), the contradiction is: $s \mid L$. In every case, we note $q_L < LM/s < M/d$.

To show that (7) holds with $n = q_L$, we proceed as follows. First we check that $[q_L/M]$ and $[q_L t/M]$ are both 0. Indeed,

$$q_L t/M = t(LM - r_L)/(Ms) = tL/s - tr_L/(Ms) < 1 - tr_L/(Ms).$$

Next we check that $[q_L s/M] < L, [q_L u/M] \geqslant L$. Indeed $q_L s/M = L - r_L/M < L$; and finally

$$q_L u/M = q_L s/M + q_L(t-1)/M \geqslant L - r_L/M + r_L/M.$$

Thus (7) holds with $n = q_L$. □

LEMMA 4. *Let B', C', D', M be positive integers with* (i) $B' \neq C'$, $B' \neq D'$, (iii) $0 < B', C', D' < M$, (iv) $1 + B' = C' + D' < M$, (vi_2) $(B', M) = (D', M) = d \geqslant 1$, (ix) $(C', M) = 1$. *Then there exists $n > 0$ (and hence infinitely many $n > 0$) such that*

$$[n/M] + [nB'/M] \neq [nC'/M] + [nD'/M]. \tag{8}$$

Proof. If all three inequalities $2B' \geqslant M, 2C' < M, 2D' < M$ are satisfied, then (8) holds with $n = 2$, so we study alternatives. If $2B' \leqslant M$, then clearly $2C' < M, 2D' < M$. We choose $u = B'$, $s = \max[C', D']$, $t = \min[C', D']$, and apply Lemma 3; the conclusion is that (7), and hence (8), holds.

The final alternative is that $2B' > M$, and in addition, either $2C' \geqslant M$ or $2D' \geqslant M$. We set $X = \max[C', D']$, $Y = \min[C', D']$, $u = M - X$, $v = M - B'$, $w = Y$. Let k be an integer not divisible by M/d. Then $M \nmid k$, and the relations

$$[k/M] + [ku/M] = [k/M] + k - 1 - [kX/M],$$
$$[kv/M] + [kw/M] = [kY/M] + k - 1 - [kB'/M]$$

hold. Thus, the relation

$$[k/M] + [ku/M] = [kv/M] + [kw/M]$$

holds if and only if the equality

$$[k/M] + [kB'/M] = [kX/M] + [kY/M]$$

is valid.

By definition, this last equality can be written

$$[k/M] + [kB'/M] = [kC'/M] + [kD'/M].$$

For the final step we set $s = \max[v, w]$, $t = \min[v, w]$, and apply Lemma 3. Here,

$$[nv/M] + [nw/M] = [ns/M] + [nt/M].$$

The conclusion of Lemma 3 is that there exists $n, 0 < n < M/d$, such that

$$[n/M] + [nu/M] \neq [nv/M] + [nw/M];$$

in fact there are infinitely many such n (they differ by a multiple of M). □

Final comments. The generalization $\Sigma[a_i n] = \Sigma[c_i n]$ seems to be somewhat awkward. In this generalization, it is still true that $\Sigma a_i = \Sigma c_i$. We take $0 \leqslant a_i, c_i < 1$. The terms on the two sides can be paired so that, e.g., $[a_1 n] + [a_2 n] = [c_1 n] + [c_2 n]$, and this seems to be essentially the only possibility. The editors have no proof.

Parts (i), (ii) or both were also solved by Merrill Barneby, Ken Brown, Alberto Cáceres (Puerto Rico), O. P. Lossers (The Netherlands), Man Kam Kwong, Ed McCravy, R. B. Macneil, Nicholas Passell, Allen J. Schwenk, George Shulman, Paul A. Vojta, David Zeitlin, and the proposer.

After the editors completed preparing the above copy, they were informed that an article, "On $(na)+(nb)=(nc)+(nd)$ for all positive integers n," by P. Noordzij had appeared in Nieuw Archief voor Wiskunde (3) 29 (1981) 59–70. Neither this MONTHLY nor E 2752 is cited there!!

Asymptotic Estimate of a Quadrature Formula

E 2868 [1981, 66]. *Proposed by Ben B. Bowen, Vallejo, California.*

Find an asymptotic estimate ($n \to \infty$, n odd) for

$$A_k(n) = \int_0^n (x-k)^{-1} \prod_{i=0}^{n} (x-i)\, dx.$$

Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands. The Newton-Coates approximation to $\int_0^1 f(x)\,dx$ is

$$Q_n(f) = \sum_{k=0}^{n} \lambda_{nk} f\left(\frac{k}{n}\right),$$

where

$$\begin{aligned}\lambda_{nk} &= \frac{(-1)^{n-k}}{k!(n-k)!}\int_0^1 nx(nx-1)\cdots(nx-k+1)(nx-k-1)\cdots(nx-n)\,dx\\ &= \frac{(-1)^{n-k}}{k!(n-k)!}\frac{1}{n}A_k(n) \qquad (1)\end{aligned}$$

(in the notation of the problem).

For the asymptotics of λ_{nk} $(n \to \infty)$ we refer, e.g., to H. Brass, *Quadraturverfahren*, Vandenhoeck and Ruprecht, Gottingen 1977 (pp. 124–125). Substitution in (1) leads to the result

$$A_k(n) = \frac{-n!}{(\log n)^2}(-1)^{n-k}\left[\frac{(-1)^k}{k} + \frac{(-1)^{n-k}}{n-k}\right]\left(1 + O\left(\frac{1}{\log n}\right)\right),$$

$$(n \to \infty, 1 \leqslant k \leqslant n-1), \quad (2)$$

$$A_0(n) = A_n(n) = \frac{(-1)^n n!}{\log n}\left[1 + O\left(\frac{1}{\log n}\right)\right], \qquad (n \to \infty). \qquad (3)$$

Since it is not easily accessible to many readers of this MONTHLY we shall give a proof of (2) following G. Pólya, *Ueber die Konvergenz von Quadraturverfahren*, Math. Zeitschrift 37 (1933) 264–286.

$$\begin{aligned}A_k(n) &= \int_0^n \frac{\Gamma(x+1)}{\Gamma(x-n)}\frac{dx}{(x-k)}\\ &= \frac{(-1)^n}{\pi}\int_0^n \Gamma(x+1)\Gamma(n+1-x)\sin \pi x \frac{dx}{(x-k)}\\ &= -\frac{1}{\pi}\left[J_{n,k} + (-1)^n J_{n,n-k}\right], \qquad (4)\end{aligned}$$

where

$$J_{n,k} = \int_0^{n/2} \Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}\,dx.$$

Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

T:	Textbook	13-18:	Grade Level
S:	Supplementary Reading	1-4 :	Time in Semesters
P:	Professional Reading	** :	Special Emphasis
L:	Undergraduate Library	?? :	Questionable

General, S. Cryptograms and Spygrams. Norma Gleason. Dover, 1981, ix + 112 pp, $3.50 (P). [ISBN: 0-486-24036-3] A collection of over one hundred "spygrams," arranged so as to introduce the most common coding techniques and to lead the novice into the subject. Deciphering tricks and clues throughout, plus appendices with tips and techniques, the 1,000 most common English words and other aids, and suggestions for further exploration. LCL

General, P. Logic and Algorithmic: An International Symposium Held in Honour of Ernst Specker. E. Engeler, H. Läuchli, V. Strassen. L'Enseignement Math, 1982, 392 pp, 90 Frs. (P). Following a brief sketch of Specker's mathematical work by Hao Wang are twenty-two contributed research or expository papers spanning a wide range of fields from algebra, combinatorics and computational complexity to model theory and proof theory. GHM

General, P. Lecture Notes in Mathematics-901: Séminaire Bourbaki Vol. 1980/81, Exposés 561-578. Springer-Verlag, 1981, 299 pp, $16.10 (P). [ISBN: 0-387-11176-X] In addition to the presentations indicated in the title, this volume contains an author index for the period 1967 to 1981. JAS

General, P?. Do You Panic About Maths? Coping with Maths Anxiety. Laurie Buxton. Heinemann Educ Books, 1981, 168 pp, $12 (P). [ISBN: 0-435-50101-1] Attempt at developing a theory of mathematics anxiety by analysis of math anxious individuals through observation of weekly group sessions and individual interviews. No new information; some of the theory seems rather contrived. MW

General, P. Mathematical Reviews Annual Index-1981. Ed: J.L. Selfridge. AMS, 1981, $138 set (P). Author Index A-L, 451 pp; Author Index M-Z, 492 pp; Subject Index 00-58, 514 pp; Subject Index 60-94, 442 pp. Also called Index of Mathematical Papers, Volume 13.

General, S(13-15), L**.** How To Read and Do Proofs: An Introduction to Mathematical Thought Process. Daniel Solow. Wiley, 1982, xiv + 172 pp, $7.95 (P). [ISBN: 0-471-86645-8] A rare, down-to-earth exposition that keeps closely in touch with the students' perspective while equipping them to understand the procedures and intentions underlying typical condensed proofs in textbooks. Develops its own systematic vocabulary for discussing strategies involving quantifiers and logical connectives as encountered in practice (not in disembodied formal logic). You may find this an invaluable supplement for that old nemesis, the "first rigorous course" in algebra or real analysis. GHM

General, S(13-16). Keno Handbook. Jim Claussen. GBC Pr, 1982, 124 pp, $4.95 (P). [ISBN: 0-89650-777-7] How-to-play book, with detailed information on filling out and figuring winnings on complicated Keno "tickets" (an interesting mathematical exercise). Plays down the disadvantageous odds and holds out the hope of winning big with statements such as "the laws of probability have no practical significance when playing Keno." RSK

General, S(13-16). Liars Poker, A Winning Strategy. John Archer. GBC Pr, 1982, 70 pp, $2.95 (P). [ISBN: 0-89650-793-9] Description of this relatively new gambling game, usually played with the eight digits of the serial number of a dollar bill. Includes conventions, strategies, variations and some mathematical analysis. RSK

General, S?(13-16). Basic Roulette. Bert Walker. GBC Pr, 1982, 80 pp, $2.95 (P). [ISBN: 0-89650-618-5] Another potentially misleading book from GBC (Gambler's Book Club) Press. Provides fundamental rules and method of play, and describes some systems which use an anti-martingale betting strategy. Uses a valid probability law to imply (but not state) that one is likely to come out ahead using these methods. RSK

Elementary, T(13: 1). College Algebra with Calculator Applications. Joseph Elich, Carletta J. Elich. Addison-Wesley, 1982, xvi + 463 pp, $19.95. [ISBN: 0-201-13340-7] Emphasizes the function idea and the use of a calculator. Considers also systems of equations and inequalities, sequences

and geometric series, discrete probability, conic sections, and the algebra of 2 x 2 matrices. FLW

Mathematics Appreciation, T(13-14: 1, 2), S. Mathematics as a Second Language, Third Edition. Joseph Newmark, Frances Lake. Addison-Wesley, 1982, xxii + 662 pp, $20.95. [ISBN: 0-201-05292-X] A rather elementary liberal arts or finite mathematics text. Few prerequisites--high school algebra and rational arithmetic are reviewed. Covers aspects of set theory, logic, algebra, functions and graphs, probability, statistics, game theory, matrices, computing, linear programming. Each chapter is introduced with a "real-life" problem, and includes explicit objectives, historical vignettes, summary, and two mastery tests. (TR, First Edition, May 1974; Second Edition, November 1977.) PZ

Precalculus, T(13: 1), S, L? Modern College Algebra. Robinson H. Parson. Vantage Pr, 1981, xii + 379 pp, $17.95. [ISBN: 533-04868-0] A fairly traditional college algebra text presented in an old fashioned style. Includes lots of examples and exercises, but not the objectives, review sections and fancy graphics of most recent books in this area. CEC

Precalculus, T(13: 1). Trigonometry: A Calculator Approach. Herman R. Hyatt, Laurence Small. Wiley, 1982, xi + 430 pp, $23.95. [ISBN: 0-471-07985-5] Integrates calculator usage in a presentation of standard right triangle trigonometry. Exercises indicate whether exact answers or calculator approximations are to be used. An appendix gives calculator instructions. Instructor's Manual and Student's Supplement are available. JNC

Precalculus, T*(13: 1). Plane Trigonometry. R. David Gustafson, Peter D. Frisk. Brooks/Cole Pub, 1982, xiii + 367 pp, $18.95. [ISBN: 0-8185-0483-8] Very nice treatment of standard material. Calculator-oriented. Attractive graphics. Trigonometric functions of angles precedes trigonometric functions of numbers. Chapter on complex numbers and polar coordinates. Readable. Interesting, if not deep, applications in text and in problems. Lots of exercises. JK

Precalculus, T(13: 1-3). Algebra and Trigonometry: A Functions Approach, Third Edition. M.L. Keedy, Marvin L. Bittinger. Addison-Wesley, 1982, xiii + 717 pp, $19.95 (P) [ISBN: 0-201-13404-7]; Trigonometry: Triangles and Functions, Third Edition. ix + 349 pp, $18.95 (P) [ISBN: 0-201-13408-X]; College Algebra: A Functions Approach, Third Edition. x + 530 pp, $18.95. [ISBN: 0-201-13400-4] Comprehensive, well organized yet flexible coverage of all major precalculus topics. The latter two texts consist of selections from the first, with some overlap. Format suitable for self-study or lecture course. Answer, student solution and test booklets available. This attractive text is worth considering for adoption. (TR, First Edition, October 1974; Second Edition, October 1978.) GHM

Precalculus, T*. Contemporary College Algebra. Jack R. Britton, Ignacio Bello. Harper & Row, 1982, x + 384 pp. [ISBN: 0-06-040987-8]; Contemporary College Algebra and Trigonometry. x + 581 pp. [ISBN: 0-06-040989-4] Well written, motivated by interesting applications, and including calculator directions and exercises, these two texts (the second adds four chapters on trigonometry to the content of the first) make the usual topics relevant to contemporary readers. However, the overuse of various blue and black inks, placement of definitions and theorems in the margin and frequent tables and boxed rules reduce some pages to distracting clutter. Instructor's Manual available. JNC

Precalculus, T*(13: 1). College Algebra. Charles Henry Brase, Corrinne Pellillo Brase. DC Heath, 1982, xvi + 457 pp, $15.95. [ISBN: 0-669-02432-5] Written with a problem solving approach in mind, the book contains numerous examples, exercises and review problems. Attractively presented, using color and appropriate figures. TAV

Precalculus, T*(13: 1), S. Precalculus with Calculator Applications. Joseph Elich, Carletta J. Elich. Addison-Wesley, 1982, xvii + 600 pp, $20.95. [ISBN: 0-201-13345-8] Covers the standard topics. The use of calculators (algebraic and RPN) is integrated into the text. Lots of good illustrations, worked examples, problems, review exercises and applications. An appealing text. CEC

Education, S(16-18), P. Mathematics for the Middle Grades (5-9): 1982 Yearbook. Linda Silvey, James R. Smart. NCTM, 1982, x + 246 pp, $13.95. [ISBN: 0-87353-192-2] Twenty-eight papers in three sections: critical issues in middle school mathematics, ideas for activities, and motivation through games and contests. Discussions of the impact on learning and teaching mathematics of sex-related differences and learning disabilities. Suggestions for incorporating computer literacy and problem solving into the curriculum. Several of the activities emphasize geometrical concepts. MW

History, P, L. Infinity and Continuity in Ancient and Medieval Thought. Ed: Norman Kretzmann. Cornell U Pr, 1982, 367 pp, $27.50. [ISBN: 0-8014-1444-X] Eleven original essays based on papers presented at the conference on Infinity, Continuity and Indivisibility in Antiquity and the Middle Ages, Cornell, April 1979. GHM

History, P, L. Norbert Wiener: Collected Works with Commentaries, Volume III. Ed: P. Masani. MIT Pr, 1981, xiii + 753 pp, $50. [ISBN: 0-262-23107-7] Third of four volumes in Wiener's Collected Works, this one containing papers on the Hopf-Wiener integral equation, on prediction and filtering, on quantum mechanics and relativity, and on miscellaneous mathematical topics. Papers are grouped by topic and introduced by commentaries linking them to contemporary research. (Volume I, TR, May 1979; Volume II, TR, April 1980.) LAS

Foundations, P. Fundamentals of Generalized Recursion Theory. Melvin Fitting. Stud. in Logic. & Found. of Math., V. 105. North-Holland, 1981, xx + 307 pp, $63.75. [ISBN: 0-444-86171-8] Generalized recursion theory (e.g., admissible set theory, recursion in higher types) treats arbitrary structures (not just integers) and broader classes of idealized "hardware" (infinite searches, etc.). The unified axiomatic treatment given here is based on generalizations of the very elegant "elementary formal systems" of Smullyan, which are simple axiomatic systems of derivation (Post production systems). GHM

Foundations, P, L. The Theory of Indistinguishables: A Search for Explanatory Principles Below the Level of Physics. A.F. Parker-Rhodes. Synthese Lib., V. 150. D Reidel Pub, 1981, xiv + 216 pp, $39.50. [ISBN: 90-277-1214-X] Develops a novel algebraic-combinatorial theory of indistinguishables, parallel to but largely unlike set theory. Explores the empirical evidence for and the philosophical consequences of the hypothesis that the physical world is a way of observing an infinite collection of indistinguishable objects. GHM

Foundations, P. Coding the Universe. A. Beller, R. Jensen, P. Welch. London Math. Soc. Lecture Note Ser., V. 47. Cambridge U Pr, 1982, 353 pp, $34.50 (P). [ISBN: 0-521-28040-4] Exposition of a major theorem of Jensen to the effect that any suitable model M of set theory can be generically extended to a "constructible" model N which contains a single set of integers which encodes all information about cardinalities and cofinalities of sets in N. GHM

Foundations, P. A Bibliography of Lambda-Calculi, Combinatory Logics and Related Topics. A. Rezus. Math Centrum, 1982, i + 86 pp, Dfl. 11,55 (P). [ISBN: 90-6196-234-X] Nearly 1000 papers and books listed alphabetically by author, with no annotation, index or subject classifications. GHM

Foundations, P. The Core Model. A. Dodd. London Math. Soc. Lect. Note Ser., V. 61. Cambridge U Pr, 1982, xxxviii + 229 pp, $24.95 (P). [ISBN: 0-521-28530-5] The core model, a generalization of Gödel's constructible universe, is used to obtain fine structural information in set theory. For example in Chapter 11 it is proved that iterable premice are acceptable, and in Chapter 12 a mouse is actually constructed. This first exposition of the subject will be welcomed by specialists in set theory. GHM

Combinatorics, S(17), P*, L*. Research Problems in Discrete Geometry. William Moser (Dept. of Math., McGill U., 805 Sherbrooke St. W., Montreal, Quebec, Canada H3A 2K6), 1981, iii + 55 pp, (P). The sixth annual edition of this collection. Includes sixty-eight problems. Each problem is stated, and the main results and a bibliography are given. A valuable source. CEC

Combinatorics, S(17), P*. Convexity and Related Combinatorial Geometry. Ed: David C. Kay, Marilyn Breen. Lect. Notes in Pure & Appl. Math., V. 76. Dekker, 1982, viii + 243 pp, $29.75 (P). [ISBN: 0-8247-1278-1] A collection of 20 original papers on convexity and combinatorics which were presented at a conference held at the University of Oklahoma, March 13-15, 1980. Includes a wide variety of research problems. CEC

Number Theory, T(13: 1), S, L. Computers in Number Theory. Donald D. Spencer. Computer Sci Pr, 1982, xi + 250 pp, $10.95 (P). [ISBN: 0-914894-27-7] Half of this book consists of an introduction to computing and programming in Basic. The other half is devoted to exploring some elementary concepts in number theory with the aid of some Basic programs. Concepts include factorization, primes, magic squares, Fibonacci numbers and modular arithmetic. CEC

Number Theory, T(14: 1), S*, L. A Pathway Into Number Theory. R.P. Burn. Cambridge U Pr, 1982, vii + 257 pp, $37.50; $14.95 (P). [ISBN: 0-521-24118-9; 0-521-28534-8] The topics in this book form a solid beginning course in elementary number theory. The approach consists of sequences of graded problems going from the specific to the general. Well thought out. Would make an excellent text for an independent study or non-lecture course. Includes solutions to the problems. CEC

Algebra, T(17-18), P. The Representation Theory of the Symmetric Group. Gordon James, Adalbert Kerber. Encyclopedia of Math., V. 16. Addison-Wesley, 1981, xxviii + 510 pp, $44.50. [ISBN: 0-201-13515-9] A general account of the ordinary and the modular representation theory of the symmetric group by two authors who have contributed much to the theory (including three Springer Lecture Notes: TR, June-July 1972; June-July 1976; August-September 1979). The use of "Young tableaux" to classify the irreducible representations of S_n gives the subject a very combinatorial flavor; in fact, the interplay between algebra and combinatorics (e.g., Polya theory) is evident throughout. Includes complete references, historical remarks, large collection of character tables. LCL

Algebra, S(17), P. Ordered Permutation Groups. A.M.W. Glass. London Math. Soc. Lect. Note Ser., V. 55. Cambridge U Pr, 1981, xlix + 266 pp, (P). [ISBN: 0-521-24190-1] A uniform, systematic account of the theory of groups of order-preserving permutations of totally ordered sets. Includes several theorems which have not previously been published, a list of unsolved problems, and an annotated bibliography. CEC

Algebra, P. Injective Modules and Injective Quotient Rings. Carl Faith. Lect. Notes in Pure & Appl. Math., V. 72. Dekker, 1982, viii + 105 pp, $19.50 (P). [ISBN: 0-8247-1632-9] Two parts: The first part concerns injective modules over Levitzki rings with a proof of the converse of the Teply-Miller-Hansen theorem. This is a sequel to Volume 49 of the Springer Lecture Notes in Mathematics with the same name. Part two is on the subject of (finitely) pseudo-Frobenius rings. JAS

Algebra, T(18: 1), S, P. Selected Topics on Polynomials. Andrzej Schinzel. U of Michigan Pr, 1982, xxi + 250 pp, $9.95 (P). [ISBN: 0-472-08026-1] An extensive collection of results on polynomials which the author has chosen on the basis of simplicity of formulation notwithstanding a certain depth. The book originated in a course of lectures given at the University of Michigan. Includes a list of references. CEC

Algebra, P. Finite Algebra and Multiple-Valued Logic. Ed: B. Csákány, I. Rosenberg. North-Holland, 1981, 880 pp, $135. [ISBN: 0-444-85439-8] Proceedings of a conference held at Szeged, Hungary, from August 27 to 31, 1979. The price seems rather steep for typescript. JAS

Algebra, T(16-17: 2), S, L. Algebra, Volume 1, Second Edition. P.M. Cohn. Wiley, 1982, xv + 410 pp, $23.95 (P). [ISBN: 0-471-10169-9] This edition includes additional material (e.g., affine spaces, linear programming, duality, the derived group, Sylow theorems, plus a complete set of answers to the exercises) and numerous improvements in exposition (expanded proofs, worked examples, etc.). (TR, First Edition, October 1974.) LCL

Algebra, T(18: 2, 3), P. Finite Groups II. B. Huppert, N. Blackburn. Grund. der math. Wissenschaften, B. 242. Springer-Verlag, 1982, xiii + 531 pp, $68. [ISBN: 0-387-10632-4] This long-awaited volume focuses on the interplay between finite group theory and linear algebra. The three lengthy chapters cover representation theory, nilpotent groups, and solvable groups. A well-documented and complete reference. SG

Finite Mathematics, T*(13: 2). Finite Mathematics and Calculus with Business, Social, and Behavioral Science Applications. Lawrence E. Spence. Harper & Row, 1982, viii + 760 pp, $24.95. [ISBN: 0-06-046368-6] A clearly written, intuitive presentation of differentiation and its applications, definite and indefinite integration, and introductory multivariable calculus has been added to material previously published in Finite Mathematics (TR, January 1982). Instructor's Manual and Student Solution Manual are available. JNC

Finite Mathematics, T(13: 2), L. Mathematics for Business, Economics, and Management. Marvin Bittinger, J. Conrad Crown. Addison-Wesley, 1982, xi + 818 pp, $24.95. [ISBN: 0-201-10104-1] A text for a one-year course covering those topics of finite mathematics and calculus most applicable to a business program. Necessarily non-rigorous, the text relies upon prior experiences and examples to motivate and introduce ideas. Clearly written and illustrated with numerous exercises. TAV

Calculus, T(13: 3), S. Calculus, Third Edition. Lynn H. Loomis. Addison-Wesley, 1982, xvi + 1022 pp, $32.95. [ISBN: 0-201-05045-5] Major changes from the Second Edition (TR, First Edition, January 1975; Second Edition, August-September 1977) include: wholly new discussions of limit, continuity, definite integral, L'Hôpital's rule, improper integrals, natural logarithm. Hand calculator exercises have been indicated. New organization provides natural break points for three term or four quarter sequence. JG

Real Analysis, T(14-15: 1, 2), S, L. Introduction to Mathematical Analysis. William R. Parzynski, Philip W. Zipse. McGraw-Hill, 1982, viii + 359 pp, $23.50. [ISBN: 0-07-048845-2] A careful reworking of one-variable calculus topics, including content and measure zero, pointwise and uniform convergence of power series. Multivariate topics include the implicit function theorem (proved using the intermediate value theorem), multiple integrals and coordinate transformations. Pedagogical features: well-chosen figures, many examples embedded in the exposition, sample continuity proofs for explicit functions. PZ

Real Analysis, S*(14-16), P, L*. A Primer of Real Functions, Third Edition. Ralph P. Boas, Jr. Carus Math. Mono., No. 13. MAA, 1981, xi + 232 pp, $16.50. [ISBN: 0-88385-000-1] Reprint, with some additional references, of the 1972 Second Edition (TR, January 1974). LAS

Complex Analysis, T(17-18: 1, 2), S, P. Several Complex Variables and Complex Manifolds. Mike Field. Cambridge U Pr, 1982. I. London Math. Soc. Lect. Note Ser., No. 65. x + 198 pp, $19.95 (P) [ISBN: 0-521-28301-9]; II. London Math. Soc. Lect. Note Ser., No. 66. vii + 211 pp, $22.95 (P). [ISBN: 0-521-28888-6] The two volumes comprise a self-contained introduction to the topic through Theorems A and B and the Kodaira embedding theorem. Volume I reviews one-variable theory, including Riemann surfaces, then develops elements of several-variable functions and analytic sets. Volume II covers calculus on complex manifolds and sheaf theory. Some references in the text are to chapters still to be written. With exercises, references to literature. PZ

Complex Analysis, P. Oswald Teichmüller: Gesammelte Abhandlungen Collected Papers. L.V. Ahlfors, F.W. Gehring. Springer-Verlag, 1982, viii + 751 pp, $43.70. [ISBN: 0-387-10899-8] Many of Teichmüller's important papers appeared only in very hard to find journals, e.g., Deutsche Mathematik. Here all 34 known papers are gathered in one place reproduced from the original journals (including some in Gothic lettering). JAS

Differential Equations, T(15-16: 1). Fourier Series and Integrals of Boundary Value Problems. J. Ray Hanna. Wiley, 1982, xi + 271 pp, $31.95. [ISBN: 0-471-08129-9] Text for an introductory course in the use of Fourier methods to solve boundary value problems: orthogonal functions, Fourier series and integrals, vibration and heat problems, Bessel functions and Legendre polynomials. Many examples and concrete exercises. LAS

Differential Equations, P. Qualitative Theory of Differential Equations. Ed: M. Farkas. North-Holland Pub, 1981. Volume I, 554 pp; Volume II, 532 pp, $159.50 set. [ISBN: 963-8021-43-8] Proceedings of the colloquium held in Szeged, Hungary in August 1979. JAS

Numerical Analysis, P. Numerical Solutions of Partial Differential Equations. Ed: John Noye. Elsevier North-Holland, 1982, xii + 647 pp, $93. [ISBN: 0-444-86356-7] Proceedings of the 1981 Conference on the Numerical Solutions of Differential Equations held at Queen's College, Melbourne University, Australia. There are review papers on finite difference techniques, Galerkin techniques, finite element methods, and boundary integral methods as well as survey articles on the numerical solution of large sparse linear algebraic systems of equations. Contributed papers discuss refinements and applications of these techniques. JAS

Numerical Analysis, T(15-16: 1, 2).** Numerical Analysis, Second Edition. Lee W. Johnson, R. Dean Riess. Addison-Wesley, 1982, xii + 563 pp, $24.95. [ISBN: 0-201-10392-3] Major changes in this edition include the addition of a chapter on partial differential equations and the inclusion of many new exercises of a computational flavor. Several of the chapters have been extensively revised and new material has been added throughout the text. (First Edition, TR, January 1978.) AO

Numerical Analysis, T(17-18: 1, 2), P, L. Computer Methods for Partial Differential Equations, Volume 1: Elliptic Equations and the Finite-Element Method. Robert Vichnevetsky. Prentice-Hall, 1981, x + 357 pp, $28.95. [ISBN: 0-13-165233-8] A "state-of-the-art" summary of numerical techniques for the solution of elliptic partial differential equations. Although the primary focus of the book is the finite element method, finite difference techniques are discussed as well. AO

Functional Analysis, P. Von Neumann Algebras. Jacques Dixmier. Math. Library, V. 27. Elsevier North-Holland, 1981, xxxviii + 437 pp, $46.75 (P). [ISBN: 0-444-86308-7] English translation of the second edition, published in French in 1969. In three parts, the first two general theory, the third special topics, with a preface on recent results by E.C. Lance. The bibliography lists about 1000 references through 1980. Has exercises. PZ

Functional Analysis, P. Approximation and Function Spaces. Ed: Zbigniew Ciesielski. Elsevier North-Holland, 1981, xiv + 897 pp, $127.75. [ISBN: 0-444-86143-2] Proceedings of the international conference held in Gdańsk, August 27-31, 1979. JAS

Analysis, P. Convex Cones. Benno Fuchssteiner, Wolfgang Lusky. Math. Stud., V. 56. Elsevier North-Holland, 1981, x + 429 pp, $46.75 (P). [ISBN: 0-444-86290-0] An outline of an elementary theory of linear functionals on convex cones. The approach develops the theory using, to an extraordinary degree, the monotonicity of functions with respect to suitable orderings rather than standard topological arguments. Measure theory plays an important role throughout the text. Many examples and applications are given and the book is quite complete, with necessary background material in the appendix. PH

Algebraic Geometry, P. Lecture Notes in Mathematics-896: Familles de Cycles Algèbriques--Schéma de Chow. Bernard Angéniol. Springer-Verlag, 1981, vi + 140 pp, $8.40 (P). [ISBN: 0-387-11169-7]

Differential Geometry, P. Seminar on Differential Geometry. Ed: Shing-Tung Yau. Annals of Math. Stud., No. 102. Princeton U Pr, 1982, ix + 706 pp, $55; $15 (P). The collected papers, except for those on closed geodesics and minimal surfaces, from the special year 1979-80 at the Institute for Advanced Study. The remaining papers have been collected in a special volume. JAS

Algebraic Topology, P. Rational Homotopy Theory and Differential Forms. Phillip A. Griffiths, John W. Morgan. Progress in Math., V. 16. Birkhauser Boston, 1981, xi + 242 pp, $14. [ISBN: 3-7643-3041-4] A monograph in which it is shown that the differential graded algebra of C^{∞} forms on a manifold can be used to calculate all of the algebraic-topological invariants of the manifold. Based on work of Sullivan. JG

Algebraic Topology, P. The Unstable Adams Spectral Sequence for Free Iterated Loop Spaces. Robert J. Wellington. Memoirs Number 258. AMS, 1982, viii + 225 pp, $12.80 (P). This monograph is concerned with computing the unstable homotopy groups of the suspensions of a topological space. JAS

Topology, P. The Branched Cyclic Coverings of 2 Bridge Knots and Links. Jerome Minkus. Memoirs Number 255. AMS, 1982, iv + 68 pp, $4.80 (P). This monograph presents the construction of a family of closed oriented 3-dimensional manifolds by pasting together pairs of regions on the boundary of a 3 ball. This generalizes lens spaces and is related to 2 bridge knots and links. JAS

Operations Research, T(15-16: 1). Essentials of Management Science. Jack R.Meredith, Efraim Turban. Business Pub, 1982, xiii + 438 pp, $13.95 (P). [ISBN: 0-256-02703-X] Intended as a short concise alternative to the comprehensive texts (e.g., Wagner or Hillier and Lieberman), the book covers lightly the typical topics of decision analysis, linear programming and variants, queues, and inventory models. The treatments are intentionally superficial, but could provide a manager some common background with operations research professionals. Problems, bibliography. TAV

Probability, P. Seminar on Stochastic Processes, 1981. Ed: E. Cinlar, K.L. Chung, R.K. Getoor. Progress in Prob. & Stat., V. 1. Birkhauser Boston, 1981, 242 pp, $16. [ISBN: 3-7643-3072-4] Eight papers presented at a three-day seminar at Evanston, Illinois in April 1981 by leading researchers in stochastic processes. TAV

Probability, T*(15: 2). Probability and Random Processes, Second Edition. S.K. Srinivasan, K.M. Mehata. Tata McGraw-Hill, 1981, x + 360 pp, Rs 27.00 (P). The first seven chapters resemble in content and style the first half of a typical post-calculus probability and statistics text, while the last two chapters cover topics on sequences of random variables and fundamental ideas on stochastic processes as mathematical entities (as opposed to a modelling approach). Exercises and examples abound. TAV

Probability, T(16: 2), P, L. Stochastic Methods in Economics and Finance. A.G. Malliaris. Advanced Textbooks in Econ., V. 17. Elsevier North-Holland, 1981, xviii + 303 pp, $35. [ISBN: 0-444-86201-3] After a rapid-fire review of relevant probabilistic concepts (e.g., Martingales, optimal stopping, and Wiener processes) and an overview of Ito's stochastic calculus with references to proofs and further developments, the author presents a large collection of applications to economics, business and finance. A useful reference for a professional or a seminar. TAV

Probability, T(15-16: 1, 2), L**.** Stochastic Processes. J. Medhi. Halsted Pr, 1982, xi + 387 pp, $17.95. [ISBN: 0-470-27000-4] Of the dozen or more texts published in the last five years aimed at the students with a background of a first course in probability and statistics but not yet to measure theory, this is the clear choice. An extremely well organized, lucidly written text with numerous problems, examples and references. The emphasis is on Markov processes with applications. TAV

Probability, T(17: 1, 2). Modern Probability Theory: An Introductory Text Book. B. Ramadas Bhat. Halsted Pr, 1981, xi + 256 pp, $14.95. [ISBN: 0-470-27039-X] While the text does not require knowledge of measure, the tone is one of measure theory in a probabilistic setting, with topics appropriate to the latter field: expectations, transforms, convergence theorems. TAV

Probability, T(17-18: 2), P*. Point Processes and Queues: Martingale Dynamics. Pierre Brémaud. Springer Series in Statistics. Springer-Verlag, 1981, xix + 354 pp, $38. [ISBN: 0-387-90536-7] A radically different approach from the usual stochastic intensity approach to point processes is used here, namely a Martingale approach which allows for a unified treatment of dynamical point process systems along the line of Wiener-driven systems. Appendices provide the necessary background and an extensive bibliography is included. TAV

Statistics, T*(13-14: 1, 2). Statistics, Second Edition. James T. McClave, Frank H. Dietrich, II. Dellen Pub, 1982, xv + 766 pp, $24.95. [ISBN: 0-89517-034-5] Revision of the 1979 First Edition. Unique features include the use of real case studies to motivate the student, and extensive coverage of regression, including multiple regression and model building. Contains over 1000 exercises, divided into two types, "learning the mechanics" and "applying the concepts." RSK

Statistics, T(17-18: 1, 2), S, P, L. Statistical Models and Methods for Lifetime Data. J.F. Lawless. Wiley, 1982, xi + 580 pp, $36.95. [ISBN: 0-471-08544-8] Estimation and inference for the appropriate parametrics models, regression models, nonparametric methods, goodness of fit tests. Some of the needed statistical topics are reviewed in an appendix. FLW

Statistics, T(13-15: 1). Elementary Statistics, Second Edition. Gene R. Sellers, Stephen A. Vardeman. Saunders Coll Pub, 1982, x + 598 pp, $24.95. [ISBN: 0-03-058456-6] Standard univariate descriptive and inferential topics, with final two chapters on least squares line and nonparametric tests. Major changes in content and organization since First Edition (TR, March 1978), including increased precision, more exercises (many based on newspaper articles), improved exposition of probability. GHM

Statistics, T(16: 1), P, L. Distribution-Free Statistical Methods. J.S. Maritz. Mono. on Appl. Prob. and Stat. Chapman & Hall, 1981, x + 265 pp, $29.95. [ISBN: 0-412-15940-6] Advanced undergraduate introduction to the ideas underlying some of the newer distribution-free methods, with emphasis on the unifying theme of randomization (or permutation). Focuses on one and two sample estimation problems, to the exclusion of many standard hypothesis test procedures. Modest number of exercises. GHM

Statistics. Mathematik heute: Grundkurs Stochastik: Lösungen und didaktisch-methodischer Kommentar. Hermann Athen, Heinz Griesel. Hermann Schroedel, 1981, 107 pp, (P). [ISBN: 3-507-83184-8] Problem solutions and teacher's commentary to be used with text of same name. JD-B

Statistics, S(15-16), P. Exercices commentés de statistique et informatique appliquées. Ronald Céhessat. Dunod, 1976, xx + 460 pp, 89 F (P). [ISBN: 2-04-011581-1]

Statistics, P. Contribution á la Théorie de la Multiestimation. Michel Lamure. Vandenhoeck & Ruprecht, 1981, 112 pp, DM 36 (P). [ISBN: 3-525-11252-1] Theoretical results concerning the estimation of parameters for multivariate distributions. LAS

Statistics, T*(14-15: 1, 2). Statistics for Business and Economics, Second Edition. James T. McClave, P. George Benson. Dellen Pub, 1982, xviii + 935 pp, $23.95. [ISBN: 0-89517-033-7] Revision of the 1978 First Edition and 1979 Revised Edition. Expanded version of Statistics, Second Edition, by McClave and Dietrich (see TR above), containing more material on graphical and numerical descriptive statistics, chapters on index numbers and time series and on survey sampling, and two chapters on decision analysis. RSK

Statistics, T(16-17: 1, 2), P*. Applied Life Data Analysis. Wayne Nelson. Wiley, 1982, xiv + 634 pp, $40.95. [ISBN: 0-471-09458-7] In the Wiley Series in Probability and Mathematical Statistics. Directed toward engineers and industrial statisticians working on product life data. Presents graphical and analytical methods for analyzing this data, which is usually censored or incomplete. Concluding chapter surveys and gives key references for topics not covered in the rest of the book. RSK

Statistics, T*(16-17: 1, 2), P. Introduction to Linear Regression Analysis. Douglas C. Montgomery, Elizabeth A. Peck. Wiley, 1982, xiii + 504 pp, $34.95. [ISBN: 0-471-05850-5] In the Wiley Series in Probability and Mathematical Statistics. Presents recent approaches as well as the standard topics, oriented toward the analyst who uses computers for problem solving. Includes many detailed real-world examples and problems. RSK

Statistics, T(17), P. Stochastic Models for Social Processes, Third Edition. D.J. Bartholomew. Wiley, 1982, xii + 365 pp, $44.95. [ISBN: 0-471-28040-2] Written for social scientists, this book is sure to place a severe strain on the mathematical prowess of most in the intended audience. Revision goals were to give a balanced view of the field (precluding simple additions to old topics), to make more usable as a text (so why no problem sets?), and to give a full bibliography. (TR, Second Edition, November 1974 and April 1982.) AWR

Statistics, T(15-16: 1, 2), S, L. Probability and Statistics for Engineering and the Sciences. Jay L. Devore. Brooks/Cole Pub, 1982, xii + 640 pp, $28.95. [ISBN: 0-8185-0514-1] Presupposes calculus. Straightforward treatment of the usual topics. More concerned with applications than derivations. Uses data from real experiments. FLW

Statistics, T(17-18: 1, 2), S, P. Aspects of Multivariate Statistical Theory. Robb J. Muirhead. Wiley, 1982, xix + 673 pp, $39.95. [ISBN: 0-471-09442-0] The statistical theory, and especially the distribution theory, behind many modern multivariate methods. FLW

Statistics, P. Identification and Informative Sample Size. H.H. Tigelaar. Math. Centre Tracts, No. 147. Math Centrum, 1982, ii + 144 pp, Dfl. 18,90 (P). [ISBN: 90-6196-235-8]

Statistics, T*(17: 1, 2). Applied Multivariate Statistical Analysis. Richard A. Johnson, Dean W. Wichern. Prentice-Hall, 1982, xiii + 594 pp, $32.95. [ISBN: 0-13-041400-X] Introductory chapters discuss prerequisite matrix algebra, simple geometry and the multivariate normal distribution. Methodological chapters treat inferences about a mean vector, comparisons of several multivariate means, multivariate linear regression models, principal components, factor analysis, discrimination and classification, and clustering. Techniques illustrated with many real data sets. RSK

Statistics, T(13-14: 1, 2). Introductory Statistics. Neil Weiss, Matthew Hassett. Addison-Wesley, 1982, xiii + 651 pp, $17.95. [ISBN: 0-201-09507-6] Text for precalculus statistics course. Covers usual topics plus chapter on planning a study. Extensive sets of examples and exercises; most use real data. Sample printouts of SPSS programs described in optional sections. Summaries and review tests at ends of chapters. KS

Computer Literacy, T?(14-15). Bits 'n Bytes About Computing: A Computer Literacy Primer. Rachelle S. Heller, C. Dianne Martin. Computer Sci Pr, 1982, x + 174 pp, $17.95. [ISBN: 0-914894-26-9] A rather fluffy outline which is possibly suitable for an elementary teacher to use as guidance in preparing materials for class use. Too shallow to provide the background to teach its own contents. Real-world students ask more substantial questions than this book is prepared to cope with. JAS

Computer Programming, L. International Microcomputer Software Directory. Ed: John Graham, Roy Wyand. Imprint Editions, 1981, xv + 402 pp, $64.95 (P). [ISBN: 0-907352-03-0] A comprehensive listing of some 5000 programs, well cross-indexed by operating system, computer, dealer, and general classification. Unfortunately this 1981 edition is rather out of date, but the publisher has made the data base available through Lockheed DIALOG. (New editions are apparently available at lower prices--consult the publisher.) JAS

Computer Programming, T(13: 1). Introduction to Programming Using Fortran 77. Glenn A. Gibson, James R. Young. Prentice-Hall, 1982, xv + 461 pp, $19.95 (P). [ISBN: 0-13-493551-9] An introduction to the Fortran 77 programming language using flowcharts and a "spiral" approach. The exercises emphasize engineering, scientific, and mathematical applications. AO

Computer Programming, S, L. Discover FORTH: Learning and Programming the FORTH Language. Thom Hogan. Osborne/McGraw-Hill, 1982, 142 pp, $15 (P). [ISBN: 0-931988-79-9] Neither a tutorial nor a programming manual, this book provides a nice introduction to the FORTH language and some of its variants. Helpful appendices and index. A few more programming examples with detailed annotations might promote better access to this rather unusual language. JAS

Computer Programming. VisiCalc: Home and Office Companion. David M. Castlewitz, Lawrence J. Chisausky, Patricia Kronberg. Osborne/McGraw-Hill, 1982, 181 pp, $15.99 (P). [ISBN: 0-931988-50-0] A number of sample spreadsheets. Includes neither a general VisiCalc manual nor consideration of how the algorithms work. JAS

Computer Science, S(15-16), L. Recursive Functions in Computer Theory. Rózsa Péter. Computers & their Appl. Ellis Horwood, 1981, 179 pp, $59.95. [ISBN: 0-85312-164-8] Exposition, largely by

examples, of some basic ideas of recursion theory as applied to problems in computer science involving formal parsing, recursive procedures, flow charts, etc. Author is well known for her contributions to the founding of recursion theory beginning in the late 1920's. This translation of the 1976 German edition is grossly overpriced for such an elementary treatment. GHM

Computer Science, T*(115-16: 1), P, L. Introduction to Microcomputers. Ed: E.L. Dagless, D. Aspinall. Digital System Design Ser. Computer Sci Pr, 1982, xi + 233 pp, $19.95. [ISBN: 0-914894-25-0] A collection of chapters giving in-depth coverage of the common ground of computer science and electronics. Sample chapters: The Computer Structure, The Instruction Set, Addressing Modes, Concurrency, Support Software, Structured Programming, and Development Environment. Applications chapters include such examples as sine and square wave generators and control of a stepper motor rather than table searching and base conversion. However, the emphasis is on intellectual content rather than "how to." JAS

Computer Science, P. Mathematical Logic in Computer Science. Ed: B. Dömölki, T. Gergely. North-Holland, 1981, 758 pp, $116.25. [ISBN: 0-444-85440-1] Thirty-one papers from a colloquium at Salgótarján, Hungary, September 1978. Bulk of the papers on semantics of programming languages, with others on program verification, applications of algebraic logic, very high level "logic programming" and methodology of language design and specification. GHM

Control Theory, P. Applications of Variational Inequalities in Stochastic Control. Alain Bensoussan, Jacques-Louis Lions. Stud. in Math. & Its Appl., V. 12. Elsevier North-Holland, 1982, xi + 564 pp, $81.50. [ISBN: 0-444-86358-3] The purpose of this book is to obtain constructive methods which can be used to calculate the solution of optimal control problems. This is done by making use of both analytic and probabilistic techniques for the solution of second order partial differential equations. AO

Systems Theory, T(16-17), P. Principles of Dynamic Programming, Part II: Advanced Theory and Applications. Robert E. Larson, John L. Casti. Control and Systems Theory, V. 7. Dekker, 1982, ix + 497 pp, $39.75. [ISBN: 0-8247-6590-7] Builds on Part I by the same authors. Requires reader background in linear algebra, probability, differential equations; some calculus of variations would also help. Lots of discussion on relief of computational complexity. AWR

Applications (Artificial Intelligence), P. Knowledge Based Theorem Proving and Learning. Donald N. Cohen. Comp. Sci. Artificial Intelligence, No. 4. UMI Research Pr, 1981, ix + 202 pp, $49.95. [ISBN: 0-8357-1202-8] Describes a large automatic theorem proving program which uses heuristics, domain specific information, a goal tree and automatic learning. Works with the typed λ-calculus. Based on a dissertation at Carnegie-Mellon University. RWN

Applications (Artificial Intelligence), P. Efficiency in Program Synthesis. Elaine Kant. Comp. Sci. Artificial Intelligence, No. 8. UMI Research Pr, 1981, ix + 162 pp, $39.95. [ISBN: 0-8357-1215-X] A revision of the author's Ph.D. thesis, this book focuses on the incorporation of efficiency considerations into automatic program synthesis systems. Both algebraic program analysis and knowledge-based techniques are used. AO

Applications (Artificial Intelligence), T(16-18: 1, 2), S, P, L. Formal Differentiation: A Program Synthesis Technique. Robert A. Paige. Comp. Sci. Artificial Intelligence, No. 6. UMI Research Pr, 1981, xi + 277 pp, $49.95. [ISBN: 0-8357-1213-3] Describes a method of program optimization in which algorithms involving high-level, concise, yet inefficient problem statements are transformed into more efficient programs. The method captures a common yet distinctive mechanism of program construction in which succinct algorithms involving costly repeated calculations are transformed into more efficient incremental versions. RJA

Applications (Physics), P. Nonlinear Phenomena in Physics and Biology. Ed: Richard H. Enns, et al. NATO Advanced Study Ser., V. 75. Plenum Pr, 1981, x + 609 pp, $75. [ISBN: 0-306-40880-5] Proceedings of the NATO Advanced Study Institute held at Banff, Alberta, Canada on August 17-29, 1980. JAS

Applications (Physics), P. Self-dual Riemannian Geometry and Instantons. Ed: Thomas Friedrich. BG Teubner, 1981, 204 pp, 19M (P). Lectures from the summer school on the Yang-Mills equation held in Kagel (East Germany) in 1979. JAS

Reviewers

RJA: Richard J. Allen, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; GHM: George H. Mills, Carleton; RWN: Richard W. Nau, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

The Mathematical Association of America
The Sixty Second Summer Meeting of the Association
University of Toronto

The Sixty Second Summer Meeting was held on the campus of the University of Toronto during the period August 20-26, 1982. There were 1335 registrants including 762 members of the Association. The Program Committee consisted of Leonard Gillman, Chairman; Edward Barbeau, Gerald Berman, John Bradburn, Richard Guy, Mabel Montgomery, Alan C. Tucker, and Daniel H. Wagner. The talks for which abstracts were submitted consisted of the following:

The Last Thirty Years of Harmonic Analysis--A Personal View, by Guido Weiss, Washington University.

The speaker described the development of harmonic analysis as he has seen it during his thirty year involvement as a researcher in this field.

Validation of the MAA BA/LC Placement Test, by William L. Drezdzon, Oakton Community College and Anita Sikes, Delgado Community College.

The BA/LC Placement Test of the MAA-PTP was sent to 49 community colleges. The test was given during the first week of class in fall 1981 to one intermediate and one college algebra class at 37 of these colleges. The results of this project show the difference in preparation.

Mathematics and Mathematicians in World War II, by J. Barkely Rosser, University of Wisconsin.

Because of the high technology of the war, schools were pressed to give special mathematical training to many people in the services. Meanwhile, hundreds of mathematicians took leave from their schools to work on special mathematical problems arising from the war effort. Ballistics, both for guns and rockets, anti-submarine tactics, aerodynamics (as plane speeds finally exceeded the speed of sound), numerical analysis (culminating in the invention of computers), quality control, coding and ciphers, problems connected with the development of the atomic bomb, and the breakthrough called Operations Research whereby mathematics could give answers to strategic and tactical problems.

Twenty Five Years of Homotopy Theory, by George W. Whitehead, Massachusetts Institute of Technology.

Hoff discovered his Eponymous Map of S3 into S2 in 1930. It was the first known example of an essential map of a sphere into a sphere of lower dimension; and its existence exemplified the strong differences between Homotopy Theory and Homology Theory. On the other hand, there are strong resemblances. The theme of this talk was how these resemblances and differences had become reconciled through the years. The talk covered the first 25 or 30 years, during which the stage was set for the great progress that has taken place in more recent times.

CUPM Recommendations for the Training of Teachers, by Donald W. Bushaw, Washington State University; John A. Dossey, Illinois State University; Marjorie M. Enneking, Portland State University; and Bruce Meserve, University of Vermont.

At intervals of about ten years, the MAA's Committee on the Undergraduate Program in Mathematics (CUPM) has published recommendations for the preservice training of mathematics teachers. Recent trends, such as the continued growth of interest in computing, and recent publications, such as NCTM's Guidelines for the Preparation of Teachers of Mathematics (1981), have made it urgent that those earlier recommendations be revised. CUPM's Teacher Training Panel has developed a draft revision of the recommendations, which were presented in outline by three members of the panel.

The Pending Death of the Mathematics Major, by Clarence F. Stephens, SUNY College at Potsdam.

Recent studies have reported a marked drop in the number of bachelor's degrees in mathematics to about 1% of all college graduates. Traditional mathematics programs are being deserted by students more interested in careers in the computing field, or to a lesser degree in other applied mathematical fields. A national panel has recommended that the traditional mathematics major be replaced by a major in the mathematical sciences. Some pedagogical approaches which have resulted in attracting a large percent of students to major in a traditional mathematics program and to elect upper division pure mathematics courses were discussed.

Some Mathematical Morsels, by Ross A. Honsberger, University of Waterloo.

A collection of elementary little problems and results, taken mostly from combinatorics and geometry. Practically none is brand new, but all were chosen to illustrate points and to be entertaining.

Recent Results in Non-Commutative Galois Theory, by M. Susan Montgomery, University of Southern California.

Galois theory for non-commutative rings was begun in 1933 by E. Noether with her work on inner automorphisms of simple algebras, and developed in the 1940's by Jacobson and Cartan in their work on division rings. These results were extended to rings of linear transformations and simple artinian rings by Azumaya and Nakayama, Hochschild, and Rosenberg and Zelinsky in the late 1940's

and 1950's. In 1978 V.K. Kharchenko continued the direction of this early work and established a Galois correspondence theorem for rings with no non-zero nilpotent ideals. Recently, Kharchenko's results have been applied to free algebras by Formanek and Dicks; also the problem of Galois intermediate rings has been considered by D.S. Passman and the speaker.

Cognition and the Classroom, by Warren Page, New York Technical College.

The perception of objects, systems, and processes vary considerably and this has tremendous bearing on how mathematical notions are perceived and utilized. Thus, teachers of mathematics must understand and appreciate individual differences in processing information if they are to find teaching strategies suited to their students' needs. Learning must be viewed as a function of many variables: culture, sex, brain lateralization, previous background experiences, and other psychophysiosocial influences. This talk attempted to experientially survey the cognitive landscape for factors which may strongly influence the learning and teaching of mathematics. It was illustrated how mathematics teachers can utilize this information in order to enhance their teaching effectiveness.

Calculus Examination and Grading By Computer, by Edward L. Spitznagel, Jr., Washington University.

Several years ago the Department of Mathematics at Washington University switched from long-answer examinations to multiple-choice, computer graded examinations in all calculus courses. Grading and record-keeping are done with the widely available statistics package SAS. Reliability and student acceptance are both higher under the new system. Automatic test generation is currently under development.

The Influence of Elasticity on Analysis, by Clifford A. Truesdell, Johns Hopkins University and Stuart Antman, University of Maryland.

Truesdell: The Classical Heritage. Elasticity was formulated and developed in the seventeenth and eighteenth centuries by mathematicians. Nonlinear problems gave rise to elliptic functions, qualitative analysis, failure of uniqueness, and the oscillating plane. Linear problems gave rise to proper frequencies, principal modes, much of the early theory of ordinary differential equations, "Bessel" and "Laguerre" functions, the concepts "function" in the modern sense and "solution" of a partial differential equation, singularities, generalized functions. Elasticity in the nineteenth century led to the "Gaussian" curvature, strain, local rotation, tensors (including properties of symmetric and orthogonal tensors and the polar decomposition), and the concept of a preferred class of solutions of partial differential equations.

Antman: Modern Developments. Nonlinear elasticity has recently stimulated some important developments in nonlinear analysis. The modern interaction between elasticity and analysis has turned out to be far more convoluted than that of the eighteenth and nineteenth centuries: certain problems of elasticity have led to new developments in analysis that are applicable to other problems of elasticity; the study of these problems has in turn resulted in further innovations in analysis. This paper traced the fascinating influence of elasticity on global bifurcation theory, the theory of variational inequalities, and the theory of quasilinear elliptic systems.

The Work of Bourbaki During the Last 30 Years, by J.A. Dieudonne, Nice, France.

What was first described was the motivation leading to the conception of Bourbaki's treatise: to establish a well-organized system of definitions and results, chosen for their efficiency as research tools in as many domains of mathematics as possible. This goal has implied extreme care in the choice of notations and terminology, and very precise formulation of these tools, in an axiomatic frame designed for optimal usefulness. The talk will examine how the mathematical community has reacted, positively or negatively, to the gradual implementation of that program during the last 30 years.

Numerical Solutions of Navier-Stokes by the Dual Variable Method, by Charles A. Hall, University of Pittsburgh.

Computational fluid dynamics is a research area which has attracted many mathematicians not only because of its importance to the engineering community, but also because of the pitfalls that are encountered in solving various discretizations of the governing Navier-Stokes equations. Such pitfalls were highlighted in this talk along with means to circumvent them.

Most two dimensional and certainly all three dimensional analyses of practical fluid flow problems tax current computer capabilities. This talk dealt with the dual variable method which uses network theory to construct matrix transformations of the discretized Navier-Stokes equations that nominally reduce by a factor of 27 the computational cost in solving two-dimensional transient fluid flow problems.

Problems of Academic Computer Centers in Four-Year Colleges, by Thomas Kurtz, Dartmouth College; Marvin L. Brubaker, Moravian College; and Burt Mendelson, Smith College.

Computer centers, whether in large universities or small colleges, usually start out as a labor of love on the part of one individual. As such, they are often associated with a department such as mathematics, at least in the minds of the faculty. Sooner or later it becomes obvious that the computer center must become "institutionalized," either because other departments begin demanding more

service or because the founder (and often the sole employee) becomes burned out. This transition is painful and expensive. The original entrepreneur is replaced by a manager, and budgets for space, equipment, and staff skyrocket. It is probably still good advice to presidents and trustees that (a) the budget for computing should be at least half the size of the budget for the book library, and (b) hardware is a continuing expense rather than a long term capital investment.

If the problems cited above are not enough, contemplate the result of the microcomputer revolution. The good news is that prices for raw computing are dropping dramatically. The bad news is that software and applications development are still expensive. This means that we can decentralize some hardware, but that on-campus, and national, networking is needed to share and communicate applications.

Some New Proofs of Some Old Theorems, by Dorothy Maharam Stone, University of Rochester.

A simple proof was given of the Birkhoff ergodic theorem that extends, almost unchanged, to a proof of the Hurewicz-Halmos ratio ergodic theorem. Proofs of other theorems were given.

The Fruitful Analogy: Analysis, Number Theory, Geometry and Topology, by Stephen S. Shatz, University of Pennsylvania.

The talk was centered on the analogy between number fields and function fields. A central theme was the interplay--suggested by complex analysis--between geometric objects associated to number theory and geometric objects over both fields of characteristic zero and characteristic p. The Zeta Function, its long history and profound influence, was a featured topic especially in view of its interpretation through topological ideas.

Also included on the program were lectures entitled "The Geometry of Some Flows in 3-Space," by John M. Franks and "Greedy Algorithym" by Alan J. Hoffman.

The Earle Raymond Hedrick Lectures: Topological, Combinatorial and Geometric Fractals

James W. Cannon of the University of Wisconsin presented the three Hedrick Lectures. The abstracts of these lectures follows:

A fractal is a set whose form is irregular or fragmented at all scales. Fractals are the consequence of iteration: any process which intersperses construction of the real numbers from the integers involves just such an iterative process, fractals are pervasive in mathematics.

Lecture I described elementary and classical fractals. Lecture II explained the role played by certain topological fractals in the characterization of topological manifolds. Lecture III examined combinatorial and geometric fractals which accompany the study of hyperbolic groups and manifolds.

Special Sessions

Mini-Courses: The Association sponsored mini-courses entitled "Uses of Computers in Undergraduate Mathematics" and "Introduction to Microprocessors in Mathematics."

Contributed Papers: For the first time at a national meeting, the Association sponsored sessions of contributed papers. There were four interest areas selected for such presentations. The interest areas and the list of presentors follows:

The Use of Computers in Undergraduate Mathematics Instruction. Presiding: Ronald H. Wenger, University of Delaware.

Session A--Experiments in Lower Division Mathematics:

"CAI in Precalculus Mathematics: A Report on the New York City Technical College CAUSE Project," by Henry Africk, New York City Technical College of the City University of New York.

"Entering and Judging Mathematical Expressions in Computer Courseware," by Morris Brooks, University of Delaware.

"Computer Assisted Learning," by Kenneth Schoen, University of Delaware.

"Intelligent Computer-Based Diagnostic Strategies and Research on Learning Topics in Precalculus," by Ronald H. Wenger, University of Delaware.

"Computer Algorithms to Select an Optimal Investment from Among Given Options," by Diane M. Spresser, James Madison University.

"Modern Techniques in Applied Mathematics Via Computers and Computer-Based Instruction," by Cliff W. Sloyer, University of Delaware.

"An Interest(ing) Problem to Solve," by Vic Norton, Bowling Green State University.

"Computer Applications of Infinite Mathematics," by Michael R. Ziegler, Marquette University.

"The Computer-Familiarization Course for the Liberal Arts Major," by Leo J. Schneider, John Carroll University.

"An Introduction to Computer Graphics," by Caren L. Diefenderfer, Hollins College.

Session B--Experiments Using Computers to Enhance Understanding of Calculus

"Computing and Calculus--A Deeper Approach," by James W. Burgmeier, University of Vermont.
"Some Non-Standard Uses of Apple Graphics," by Ron Sandstrom, Fort Hays State University.
"Computer Assisted Calculus at Northeastern," by Mark Bridger, Northeastern University.
"Calculus Using Computers and Programmable Calculators," by Donald R. Snow, Brigham Young University.
"Learning Multivariate Calculus with the Help of the Computer," by J.M.A. Danby, North Carolina State University at Raleigh.
"Using APPLE Generated Graphics in the Calculus Classroom," by Robert S. Fisk, Colorado School of Mines.
"Calculus Through the Eye of the Computer," by Sheldon P. Gordon, Suffolk County Community College.

Session C--Experiments with Computers in Upperdivision Mathematics Courses:

"Pulse Process Models for Use in Undergraduate Math Courses," by E.L. Perry, University of Houston at Clear Lake City.
"Mathematics for Computer Graphics," by Joan Wyzkoski, Bradley University.
"A Classroom Computer Applied to Y = -KY," by D.W. Blackett, Boston University.
"The Heat Equation on a Metal Bar With a Radiating End," by Howard Penn, U.S. Naval Academy.
"A Constant Surprise in Computational Mathematics," by Bruce H. Edwards, University of Florida.
"The Mathematical Foundations of Computer Graphics: An Undergraduate Course in Applied Linear Algebra," by Gerald J. Porter, University of Pennsylvania.
"Picturing Complex Integrals," by Bart Braden, Northern Kentucky University.
"Learning Linear Algebra Through APL," by Murray Eisenberg, University of Massachusetts at Amherst.

Classroom Notes. Presiding: JoAnne Growney, Bloomsburg State College.

"A Desargues´ Theorem for Desargues´ Configurations," by Andrew P. Guinand, Trent University.
"The Absolute Determinant and Integration," by J.E. LeBel, Toronto University.
"An Alternate Method for Finding the Partial Fraction Decomposition of a Rational Function," by H. Joseph Straight and Richard Dowds, SUNY College at Fredonia.
"The Chain Rule," by Cynthia P. Yang, Miami University.
"Integral Vectors," by Joyce Williams, University of Lowell.
"An Extension of the Pythagorean Theorem and a Spiral Development of a Rigorous Proof," by H. Keith Stumpff, Central Missouri State University.
"The Number of Homomorphisms from Zm Into Zn," by Joseph A. Gallian, University of Minnesota at Duluth.
"Use of Rubik´s Cube in Teaching Undergraduate Abstract Algebra," by Eliose Hamann, Elmhurst College.
"Rubik´s Cube and a Minimal Counterexample in Group Theory," by Arnold D. Feldman, Franklin and Marshall College.
"A Knockout Tournament Problem," by Curtis Cooper, Central Missouri State University.
"Four Mathematical Recreations Using Only Elementary Algebra," by Michael W. Ecker, Pennsylvania State University.

The Undergraduate Mathematics Curriculum. Presiding: Stephen B. Maurer, Swarthmore College.

"Give Us More Visual Mathematics," by Martin E. Flashman, Humboldt State University.
"A Two-Track Concentration in Applied Mathematics," by William Parzynski, Montclair State College.
"Decomposition of a Matrix Into Rank-Determining Submatrices," by Albert W. Tucker, Princeton University.
"Reflections of a Past Maverick Chairman," by Joerg Mayer, Lebanon Valley College.
"Contract Majors in Mathematics," by Lynn A. Steen, St. Olaf College.
"Harvard´s New Quantitative Reasoning Requirement," by Andrew Gleason, Harvard University.

Special Concerns: Remediation, Articulation, and Math Anxiety. Presiding: John A. Dossey, Illinois State University.

"A Multiple Regression Approach to Placement in Mathematics at the United States Coast Guard Academy," by Ernest J. Manfred, U.S. Coast Guard Academy and Robert L. Heiny, University of Northern Colorado.
"Mathematics Anxiety Among Minorities," by Queen E. Wiggs, University of the District of Columbia.
"Developmental Mathematics and Mathophobia: An Approach to Collegiate Remediation," by Irving Anellis, McMaster University.
"Cognitive-Oriented Supplementary Materials and Students´ Cognitive Processes and Performance in College Remedial Algebra," by Joseph Witkowski, University of Georgia.
"A Key Goal for Remediation--Learn to Use Mathematics," by JoAnne Growney, Bloomsburg State College.
"Strategic Methods of Mathematics Pedagogy," by Marion Ben-Jacob, Mercy College.
"Why Can´t ´Remedial Education´ Be Translated Into French, German, ...?" by Marguerite Gravez, Pennsylvania State University, Allentown.

The Local Arrangements Committee consisted of Stephen J. Pierce, Chairman; Morton Abramson; James G. Arthur; Raymond G. Ayoub; Edward J. Barbeau, Jr., Publicity Director; L. Terrell Gardner; William J. LeVeque; James McCool; David P. Roselle; James R. Vanstone; John B. Wilker.

Board and Business Meeting

The Board of Governors met at 9:00 A.M. on Sunday, August 22, 1982 in Simcoe Hall. The major items of business transacted at the meeting will be announced to the membership in FOCUS.

The Business Meeting of the Association was held at 4:30 P.M. on Wednesday, August 25, 1982 and the Carl B. Allendoerfer and George Polya Awards for expository articles in Mathematics Magazine and the Two-Year College Mathematics Journal were presented.

Election of Members

At its meeting on August 22, 1982, the Board elected to membership 580 applicants for individual membership and 5 applicants for academic membership. The latter follows:

Florida Atlantic University, Boca Raton, Florida
Hamline University, St. Paul, Minnesota
New England College, Henniker, New Hampshire
Paul Smith's College, Paul Smith's, New York
Rochester Institute of Technology, Rochester, New York

Respectfully submitted,

David P. Roselle, Secretary

C E N T E R S E C T I O N I N D E X

Volume 99, 1982

Parts (i), (ii) or both were also solved by Merrill Barneby, Ken Brown, Alberto Cáceres (Puerto Rico), O. P. Lossers (The Netherlands), Man Kam Kwong, Ed McCravy, R. B. Macneil, Nicholas Passell, Allen J. Schwenk, George Shulman, Paul A. Vojta, David Zeitlin, and the proposer.

After the editors completed preparing the above copy, they were informed that an article, "On $(na)+(nb)=(nc)+(nd)$ for all positive integers n," by P. Noordzij had appeared in Nieuw Archief voor Wiskunde (3) 29 (1981) 59–70. Neither this MONTHLY nor E 2752 is cited there!!

Asymptotic Estimate of a Quadrature Formula

E 2868 [1981, 66]. *Proposed by Ben B. Bowen, Vallejo, California.*

Find an asymptotic estimate ($n \to \infty$, n odd) for

$$A_k(n) = \int_0^n (x-k)^{-1} \prod_{i=0}^{n} (x-i)\,dx.$$

Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands. The Newton-Coates approximation to $\int_0^1 f(x)\,dx$ is

$$Q_n(f) = \sum_{k=0}^{n} \lambda_{nk} f\left(\frac{k}{n}\right),$$

where

$$\begin{aligned}\lambda_{nk} &= \frac{(-1)^{n-k}}{k!(n-k)!}\int_0^1 nx(nx-1)\cdots(nx-k+1)(nx-k-1)\cdots(nx-n)\,dx\\ &= \frac{(-1)^{n-k}}{k!(n-k)!}\frac{1}{n}A_k(n) \qquad (1)\end{aligned}$$

(in the notation of the problem).

For the asymptotics of λ_{nk} $(n\to\infty)$ we refer, e.g., to H. Brass, *Quadraturverfahren*, Vandenhoeck and Ruprecht, Gottingen 1977 (pp. 124–125). Substitution in (1) leads to the result

$$A_k(n) = \frac{-n!}{(\log n)^2}(-1)^{n-k}\left[\frac{(-1)^k}{k} + \frac{(-1)^{n-k}}{n-k}\right]\left(1 + O\left(\frac{1}{\log n}\right)\right), \qquad (n\to\infty, 1 \leqslant k \leqslant n-1), \quad (2)$$

$$A_0(n) = A_n(n) = \frac{(-1)^n n!}{\log n}\left[1 + O\left(\frac{1}{\log n}\right)\right], \qquad (n \to \infty). \quad (3)$$

Since it is not easily accessible to many readers of this MONTHLY we shall give a proof of (2) following G. Pólya, *Ueber die Konvergenz von Quadraturverfahren*, Math. Zeitschrift 37 (1933) 264–286.

$$\begin{aligned}A_k(n) &= \int_0^n \frac{\Gamma(x+1)}{\Gamma(x-n)}\frac{dx}{(x-k)}\\ &= \frac{(-1)^n}{\pi}\int_0^n \Gamma(x+1)\Gamma(n+1-x)\sin \pi x \frac{dx}{(x-k)}\\ &= -\frac{1}{\pi}\left[J_{n,k} + (-1)^n J_{n,n-k}\right], \qquad (4)\end{aligned}$$

where

$$J_{n,k} = \int_0^{n/2}\Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}dx.$$

In the following we use the fact that $|(\sin x)/x| \leqslant 1$. We also need that $\Gamma(x+1)\Gamma(n+1-x)$ is decreasing for $0 < x < n/2$. To show this we use the well-known Beta-function integral

$$\frac{\Gamma(x+1)\Gamma(n+1-x)}{\Gamma(2+n)} = \int_0^1 t^x(1-t)^{n-x}\,dt$$

$$= \tfrac{1}{2}\int_0^1 [t(1-t)]^{n/2}\left\{\left(\frac{t}{1-t}\right)^{(n/2)-x} + \left(\frac{1-t}{t}\right)^{(n/2)-x}\right\} dt.$$

The result follows from the fact that (for $a > 0$) the function $a^x + a^{-x}$ is increasing on $[0, \infty)$, $(a \neq 1)$.

For the sake of simplicity we shall assume that n is even and $k \geqslant 4$; the result is also true for odd n and $k \geqslant 1$. We estimate the integral from 1 to $n/2$ by dividing this into $[1,3]$ and $[3, n/2]$. We then have

$$\left|\int_1^{n/2}\Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}dx\right| < \frac{2\Gamma(n)}{k-3} + \frac{n}{2}\Gamma(4)\Gamma(n-2)\pi,$$

and hence we find for $n \to \infty$

$$J_{n,k} = \int_0^1 \Gamma(x+1)\Gamma(n+1-x)\frac{\sin \pi x}{k-x}dx + O(k^{-1}\Gamma(n))$$

uniformly in k for $n \to \infty$. Applying Stirling's formula then yields

$$\frac{J_{n,k}}{n!} = \int_0^1 n^{-x}\left\{1 + O\left(\frac{1}{n}\right)\right\}\frac{\Gamma(x+1)\sin \pi x}{k-x}dx + O\left(\frac{1}{nk}\right).$$

Substitute $t = x\log n$. We find

$$\frac{J_{n,k}}{n!} = \frac{\pi}{k(\log n)^2}\int_0^{\log n} te^{-t}\cdot\Gamma\left(1+\frac{t}{\log n}\right)\cdot\frac{\log n}{\pi t}\sin\frac{\pi t}{\log n}\left(1-\frac{t}{k\log n}\right)^{-1}dt + O\left(\frac{1}{nk}\right).$$

The final step was the following standard argument. Let $f(t) = 1 + O(t)$, $(t \to 0)$ and $0 \leqslant f(t) \leqslant C$ on $[0,1]$. Then

$$\left|\int_0^m te^{-t}f\left(\frac{t}{m}\right)dt - \int_0^\infty te^{-t}\,dt\right| \leqslant \int_0^{2\log m} te^{-t}O\left(\frac{t}{m}\right)dt + \int_{2\log m}^m te^{-t}O(1)\,dt + \int_m^\infty te^{-t}\,dt$$

$$= O(m^{-1}) + O(m^{-2}\log m) + O(me^{-m}) = O(m^{-1}), \qquad (m \to \infty).$$

Since $f_k(t) = \Gamma(1+t)\cdot\dfrac{\sin \pi t}{\pi t}\cdot\left(1-\dfrac{t}{k}\right)^{-1}$ satisfies these conditions (also for $k=1$, in fact $f_k(t) \leqslant 2$ on $[0,1]$ for all k), we have, uniformly in k for $1 \leqslant k \leqslant n-1$,

$$\frac{J_{n,k}}{n!} = \frac{\pi}{k(\log n)^2}\left(1 + O\left(\frac{1}{\log n}\right)\right), \qquad (n \to \infty). \tag{5}$$

The proof for $k = 0$ is similar. The result (2) follows from (4) and (5).

Paul F. Byrd points out that a detailed derivation is given in V. I. Krylov, *Approximate Calculation of Integrals*, The Macmillan Co., pages 84–86.

ANSWERS TO "PHOTOS" ON PAGE 735

Top: William Feller; bottom: Ernst Zermelo.

The Diophantine Equation $x^2 + 15^a = 2^b$

E 2880 [1981, 291]. *Proposed by Leo J. Alex, SUNY, College at Oneonta.*

Find all solutions to the equation $x^2 + 15^a = 2^b$ in integers x, a, and b.

I. *Solution by Lorraine L. Foster, California State University, Northridge.* Suppose either (i) $N \equiv 1 \bmod 4$, or (ii) $N > 0$, $N \equiv 3 \bmod 4$, and N has a prime factor $p \equiv 3$ or $5 \bmod 8$. Then the equation $x^2 + N^a = 2^b$ has no solution with $b \geqslant 2\log_2 N + 2$, so that the equation has only finitely many solutions. In particular if $N = 15$, the only solutions are $(x, a, b) = (0,0,0), (\pm 1, 0, 1), (\pm 1, 1, 4), (\pm 7, 1, 6)$.

Proof. If $N \equiv 1 \pmod 4$ (or $a = 0$) clearly the only solutions are $(0,0,0), (\pm 1, 0, 1)$. Hence suppose $N \equiv 3 \pmod 4$ as above so that $(2/p) = -1$ for some $p \mid N$. Also let $b \geqslant 2\log_2 N$ and suppose (x, a, b) is a solution, $x > 0$. Using modulus 4, a, x are odd. Thus, using modulus p, $2 \mid b$, $b = 2B$. Thus $(2^B + x)(2^B - x) = N^a$ and since $(2^B + x, 2^B - x) = 1$ we have $2^B + x = R^a, 2^B - x = S^a$ $(R, S > 0, RS = N, (R, S) = 1)$. Hence $2^{B+1} = R^a + S^a, (RS^{-1})^a + 1 \equiv 0 \pmod{2^{B+1}}$. Thus the order of RS^{-1}, a power of 2, $\bmod 2^{B+1}$ divides $2a$ and hence divides 2. Thus $(RS^{-1})^a \equiv RS^{-1}$ so that $R + S \equiv 0 \pmod{2^{B+1}}$. However, $0 < R + S < 2N < 2^{[\log_2 N]+1} \leqslant 2^{B+1}$, a contradiction.

In the case $N = 15$ (with $a \neq 0, 2 < b < 2\log_2 15 < 8$, b even), we find the additional solutions given above by trial.

II. *Solution by David Leep, University of Chicago. Theorem.* Let $N > 1$ be an odd integer and assume N contains a prime factor p such that $(2/p) = -1$ (2 is not a quadratic residue). Then all integer solutions can be computed by factoring (as shown below). Moreover, all solution sets have $a \leqslant 1$.

Proof. If $b = 0$, then $b = a = 0$. If $b = 1$, then $a = 0$ and $x = \pm 1$. Now assume $b \geqslant 2$. Since x is odd, then a is odd, because $N^a \equiv 3 \bmod 4$. Set $a = 2c + 1$. Since $a \geqslant 1$, the relation $x^2 + N^a = 2^b \bmod p$ implies that b is even. Set $b = 2d$. Note $N^a = (2^d + x)(2^d - x)$. Clearly $(2^d + x, 2^d - x) = 1$. Thus $2^d + x = N_1^a, 2^d - x = N_2^a$, with $N = N_1 N_2, N_1 > N_2, (N_1, N_2) = 1$. Note that $2 \cdot 2^d = N_1^a + N_2^a = (N_1 + N_2)(N_1^{2c} + \cdots + N_2^{2c})$. The final factor is odd, so must be unity. Either $c \geqslant 1$ or $c = 0$. If $c \geqslant 1$, then $N_1^{2c-1}(N_1 - N_2) + \cdots + N_2^{2c} > 1$, a contradiction. Thus $c = 0, a = 1, 2^{d+1} = N_1 + N_2$.

Summarizing, each decomposition $N = N_1 N_2$ with $(N_1, N_2) = 1$, $N_1 > N_2$, and with $N_1 + N_2$ a power of 2, $N_1 + N_2 = 2^{d+1}$, determines d. Taking $b = 2d$ gives $x = N_1 - 2^d = 2^d - N_2, (\pm x)^2 + N = 2^b$. $(x^2 + N = (N_1 - 2^d)(2^d - N_2) + N_1 N_2 = (N_1 + N_2)2^d - 2^{2d} = 2^{2d+1} - 2^{2d} = 2^{2d} = 2^b$.)

Corollary: If $N = 15, 5 \mid 15, (2/5) = -1$, the solutions come from $N_1 = 5, N_2 = 3$ or from $N_1 = 15$, $N_2 = 1$. This gives all nontrivial solutions.

Also solved by M. Bencze (Romania), A. J. Berlau, J. Binz (Switzerland), J. A. Brandler, R. Breusch, B. Cheng & Dinh Thế Hùng (students), G. Fisher, N. Franceschine III, R. Gilmer, J. D. Griffin, E. Grosswald, W. V. Grounds, L. Harkleroad, S. F. Henderson, K. Heuer & G. A. Heuer, W. Johnson, J. A. Jolley, J. S. Lew, O. P. Lossers (The Netherlands), H. M. Marston, M. R. Modak & S. A. Katre (India), D. E. Orr, Thu Pham, R. E. Shafer, G. Shulman, S. Singh, University of South Alabama Problem Group, J. Suck (Germany), E. Tröst (Switzerland), C. Vanden Eynden, and the proposer.

Sum Involving Legendre Symbol

E 2882 [1981, 291]. *Proposed by Ronald J. Evans, University of California at San Diego.*

Let p be a prime congruent to any one of $5, 7, 11, 13, 23$ modulo 24. Write (n/p) for the Legendre symbol. Show that $\Sigma n(n/p) = 0$, if the sum is extended over all values of n congruent to $1 \bmod 6$ in the range $0 < n < 6p$.

Solution by the proposer. Write $\chi(n)=(n/p)$ and let S denote the sum in question.

Case 1. $p\equiv 5$ or $13 \pmod{24}$.
Here $\chi(-1)=1, \chi(2)=-1$. Let $p\equiv\varepsilon \pmod 6$ with $\varepsilon=+1$ or -1. Then

$$S=0 \quad\text{iff}\quad 0=\sum_{-p\varepsilon/6<k<p-p\varepsilon/6}(6k+p\varepsilon)\chi(6k+p\varepsilon) \quad\text{iff}\quad T=0,$$

where $T=\sum_{-p\varepsilon/6<k<p-p\varepsilon/6}k\chi(k)$. We have

$$T=\sum_{-p\varepsilon/6<k<p/2}k\chi(k)+\sum_{-p/2<k<-p\varepsilon/6}(k+p)\chi(k+p)=T_1+p\chi(-1)T_2,$$

where $T_1=\sum_{-p/2<k<p/2}k\chi(k)$, $T_2=\sum_{p\varepsilon/6<k<p/2}\chi(k)$. But pairs of terms in T_1 annul one another since $\chi(-1)=1$. Thus $T_1=0$. As for T_2, note that

$$\sum_{-p/6<k<p/2}\chi(k)=-\sum_{-p/2<k<-p/6}\chi(k)=-\chi(-1)T_3,$$

where $T_3=\sum_{+p/6<k<p/2}\chi(k)$. To show that $T_3=0$, we have

$$\begin{aligned}T_3&=\sum_{p/6<k<p/3}\chi(k)+\sum_{p/6<k<p/4}\chi(2k)+\sum_{p/4<k<p/3}\chi(p-2k)\\&=\sum_{p/6<k<p/3}\chi(k)+\chi(2)\Big\{\sum_{p/6<k<p/4}\chi(k)+\chi(-1)\sum_{p/4<k<p/3}\chi(k)\Big\}.\end{aligned}$$

Since $\chi(2)=-1$, T_3 must be 0. Thus $T_2=0$.

Case 2. $p\equiv 7, 11$, or $23 \pmod{24}$.
Here $\chi(-1)=-1$ and $\chi(a)=1$, where $a=2$ if $p\equiv 7$ or $23 \pmod{24}$ and $a=3$ if $p\equiv 11 \pmod{24}$. Write $b=6/a$. We note that

$$\begin{aligned}\sum_{\substack{0<m<6p\\ a\mid m,\, b\nmid m}}m\chi(m)&=a\sum_{\substack{0<k<bp\\ b\nmid k}}k\chi(k)=\sum_{\substack{0<k<bp\\ b\nmid k}}\chi(k)\sum_{j=1}^{a}(k+jbp)\\&=\sum_{j=1}^{a}\sum_{\substack{0<k<bp\\ b\nmid k}}(k+jbp)\chi(k+jbp)=\sum_{\substack{0<m<6p\\ b\nmid m}}m\chi(m).\end{aligned}$$

Subtracting the first sum from the last sum, we see that

$$V=\sum_{\substack{0<m<6p\\ 2\nmid m,\, 3\nmid m}}m\chi(m)=0.$$

Now, $V=V_1+V_5$, where

$$V_c=\sum_{\substack{0<n<6p\\ n\equiv c \pmod 6}}n\chi(n).$$

Replacing n by $6p-n$, and using $\chi(-1)=-1$, we see that $V_1=V_5$. Thus $V_1=V/2=0$.

Also solved by D. T. Hung & B. Cheng (students), L. E. Mattics, and D. Orr.

For related results (with 6 replaced by other numbers) see equation (2.6) in the proposer's paper *Pure Gauss Sums over Finite Fields*, to appear in Mathematika.

Sierpiński's Problem $x(x+1)=k^2y(y+1)$

E 2891 [1981, 444]. *Proposed by Abou-Dardaye Barry, Gabon, Central Africa.*

In "250 problems in elementary number theory," problems 140, 141, Sierpiński shows that the Diophantine equation (*) $x(x+1)=k^2y(y+1)$ has no solution with $xy\neq 0$ if $k=2$ or if $k=p^n$. Find other values of k that give the same conclusion.

Solution by J. Suck, Essen, Germany. The equation $x(x+1)=k^2y(y+1)$ has no solution in positive integers x, y if $k=mp^n>6$ where p is a prime, $n\in N$ and $m=3$ or 4, e.g., $k=24$.

Proof. Assume the contrary, that is, take $k=3p^n$ or $4p^n$ $(k>6)$.
Since

$$x(x+1) = k^2y(y+1) \Leftrightarrow (ky+x)(ky-x) = x - k^2y,$$

a solution with $ky - x \geqslant 0$ would imply that

$$ky - x + x - k^2y = ky(1-k) \geqslant 0.$$

Thus for a solution of (*), x is confined to $x = ky + b$ with $b = 1,2,\ldots$ leading to (**) $b(b+1) = k(k-(2b+1))y$ for a solution. We now check cases.

(i) If $b = cp^n$, $c \geqslant 2$, then $2b+1 \geqslant k$, contradicting $b(b+1) > 0$ in (**).
(ii) If $b = p^n$ then

$$p^n + 1\,[= m(mp^n - 2p^n - 1)y] = \big((m^2-2m)p^n - m\big)y$$

and a fortiori (since $y \geqslant 1$), $m + 1 \geqslant (m^2 - 2m - 1)p^n$ whence $m = 3$, $p = 2$, $n = 1$, contradicting $k > 6$.
(iii) If $b + 1 = cp^n$, $c > 2$, then $2b + 1 \geqslant k$, but $b(b+1) > 0$.
(iv) If $b + 1 = 2p^n$, then $2b + 1 \geqslant k - 1$, hence $2(b+1) = k$ and so $m = 4$, but then $2p^n - 1 = 2y$.
(v) If $b + 1 = p^n$, then $p^n - 1 = m(mp^n - (2p^n - 1))y \geqslant (m^2 - 2m)p^n + m$, but this is impossible for $m \geqslant 3$.

Remarks.

Technical: the problem does not specify positive x, y but for the k given above there are no others either (except for the trivial $x = y = -1$) as may be seen from turning to $-x-1$ or $-y-1$.

Substantive: the result does not hold for $m = 5$; take $k = 35$, then $x = 49$, $y = 1$ is a solution, nor does it hold for $m = 2$ unless $p = 2$ because more generally we have the solution $[x = 4y(y+1), y]$ if $k = 2(2y+1)$. Note that for any given y there are infinitely many k admitting a solution (x, y) (just let Wallis and Lord Brouncker solve $z^2 - 4y(y+1)\cdot k^2 = 1$ and set $x = (z-1)/2$). I went through the remaining $k \leqslant 103$; of these only $k = 99$ appears to admit a solution, $x = 242$, $y = 2$. I encountered no k with more than one solution; is this obvious?

Sakmar found that there is no solution if $k = 3,4,5,7,8,9$. Fisher noted $8\cdot 9 = k^2\cdot 1\cdot 2$ if $k = 6$, $24\cdot 25 = k^2\cdot 2\cdot 3$ if $k = 10$, $48\cdot 49 = k^2\cdot 3\cdot 4$ if $k = 14$. Bernstein, Grosswald, and Nelsen noted that if $k-1$, $k+1$ are both primes, there is no solution. Shafer found some of these values, as well as others. The proposer found still other values, including $k = 16p^{2n}$.

Also solved by O. P. Lossers (Netherlands).

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A 5B9, by April 30, 1983. The solver's full post-office address should be on each sheet.

6409. *Proposed by Dilip Kumar Bayen and S. K. Chatterjea, Calcutta University, India.*

Let f: $K^n \to E^n$ be a continuous function (where E^n is Euclidean n-space and $K^n = \{x \mid x \in E^n, \|x\| \leqslant 1\}$) such that for every $y \in S^{n-1} = \{x \mid x \in E^n, \|x\| = 1\}$ there is $m > 1$ with $f(y) = my$. Show that f has a fixed point. Compare this MONTHLY, 78 (1971) 310, #5721.

6410. *Proposed by Gary Gundersen, University of New Orleans, and Steve Osborn, Dallas, Texas.*

Let $f(z)$ and $g(z)$ be two nonconstant rational functions such that if $g(z_0)$ is an integer for

some z_0, then $f(z_0)$ is also an integer. Show that $f(z) = P(g(z))$ where P is a polynomial with rational coefficients.

6411. *Proposed by Leo J. Alex, SUNY Oneonta.*

Find all solutions in integers a, b, c, d to the equation $1 + 3^a = 5^b + 3^c 5^d$.

6412. *Proposed by F. S. Cater, Portland State University.*

Let $f: \mathbb{R} \to \mathbb{R}$ be a function and let $G = \{(x, f(x)): x \in \mathbb{R}\}$ be its graph. Prove the following:

(i) If f is not continuous everywhere and G is connected, then $\overline{G}$ contains a line segment which is perpendicular to the x-axis. Thus G is not closed.
(ii) There exists a function f whose graph G is a connected and dense subset of $\mathbb{R}^2$. This f is nowhere continuous.
(iii) There exists a function f which is continuous at every irrational point and discontinuous at every rational point and whose graph G is connected.

6413. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

The positive integers are partitioned into two sets A, B so that $A + A$ and $B + B$ each represent no more than finitely many primes. Show that the partition is unique. (This generalizes E 2977.)

6414. *Proposed by A. Wilansky, Lehigh University.*

Suppose the matrix $A = (a_{nk})$ transforms the sequence $x = \{x_n\}$ into the sequence $y = \{y_n\}$, so that

$$y_n = \sum_{k=1}^{\infty} a_{nk} x_k \quad \text{for } n = 1, 2, \ldots .$$

If

$$y_n = \begin{cases} x_1 & n \text{ even,} \\ x_{(n+1)/2} & n \text{ odd,} \end{cases}$$

observe that (i) $a_{nk} = 0$ for $k > n$; (ii) x bounded implies y bounded; (iii) y convergent implies x convergent; (iv) there is a convergent x for which y is not convergent.

Is there a matrix A with properties (i), (ii), (iii), and (iv) for which $a_{nn} \neq 0$ for all n?

SOLUTIONS OF ADVANCED PROBLEMS

Extensions of Functions on Metric Spaces

6289 [1980, 140]. *Proposed by James W. Fickett, University of Colorado, Boulder.*

Let A be a subset of the metric space M. Show that the two following conditions are equivalent.

(1) A is a G_δ in M.

(2) Any continuous function f mapping A into another metric space N has an extension F: $M \to N$ which is continuous at each point of A (considered as a point of M).

Composite of solutions by Robert B. Israel, University of British Columbia, and the proposer.

(1) $\to$ (2): It suffices to extend the identity function on A to a function r: $M \to A$ which is continuous at each point of A. For any f as in (2) we may then put $F = f \circ r$.

Let $F_1 \subset F_2 \subset F_3 \subset \cdots$ be closed subsets of M with F_n^c contained in the $(1/n)$-neighborhood

of A for each $n > 1$, and with $\cup_{n=1}^{\infty} F_n = A^c$. Then it is easy to see that the r: $M \to A$ defined as follows has the desired properties.

$$r(x) = \begin{cases} x & \text{if } x \in A, \\ \text{any } y \in A & \text{if } x \in F_1, \\ \text{any } y \in A \text{ with } d_M(x, y) < 1/(n-1) & \text{if } x \in F_n \setminus F_{n-1}, n > 1. \end{cases}$$

(2) $\to$ (1): Let $N = A$ and f: $A \to A$ be the identity function. Let F: $M \to A$ be the extension of f given by (2) and define $K_n = \{x \in M\colon\ d_M(F(x), x) < 1/n\}$. Then each K_n contains a neighborhood of each point of A (because F is continuous there), and $\cap_{n=1}^{\infty} K_n = \{x \in M: F(x) = x\} = A$. Thus A is a G_δ.

Also solved by F. S. Cater, Portland State University; Chico Problem Group, California State University at Chico; James N. Hagler, The Catholic University of America; and John C. Tripp, Southeast Missouri State University.

Subfields of $\mathbb{C}$ of index 2

6310 [1980, 675]. *Proposed by Seth Warner, Duke University.*

Is there a subfield K of the complex number field $\mathbb{C}$ such that $\mathbb{C}$ is two-dimensional over K and
(a) K is isomorphic to a proper subfield of the real number field $\mathbb{R}$?
(b) K is not isomorphic to any subfield of $\mathbb{R}$?

Solution by Daniel B. Shapiro, Ohio State University. (a) Yes. We rely on the theorem: Let A be a transcendence basis of $\mathbb{R}/\mathbb{Q}$, and set $F = \mathbb{Q}(\{\alpha | \alpha \in A\})$. Then $\mathbb{C}$ is the algebraic closure of F.

Let $A' \subset A$ be a proper subset with the same cardinality as A, and let $F' = \mathbb{Q}(\{\alpha | \alpha \in A'\})$ (field adjunction). Take $K' = \{\beta \in \mathbb{R} | \beta$ is algebraic over $F'\}$, the algebraic closure of F' in $\mathbb{R}$. Then K' is real closed and K' is a proper subfield of $\mathbb{R}$. Also, $K'(i) \equiv K'(\sqrt{-1})$ is algebraically closed, since K' is real closed. Thus $K'(i) \cong$ an algebraic closure of F'. By uniqueness of algebraic closures, using $F \cong F'$, it follows that $K'(i) \cong \mathbb{C}$. Let K be the image of K' under this isomorphism. Then $[\mathbb{C} : K] = 2$, and K is isomorphic to the proper subfield K' of $\mathbb{R}$.

(b) Yes. Choose a nonarchimedean ordering $\prec$ on F; this is easy to do for a purely transcendental extension of $\mathbb{Q}$. Let K be a real closure in $\mathbb{C}$ of the ordered field $(F, \prec)$. Then $[C: K] = 2$, but K cannot be embedded in $\mathbb{R}$, since the (unique) ordering on K is nonarchimedean.

E. R. Gentile (Argentina) took t transcendental, $L = \mathbb{R}(t)$, $t > q (\forall q \in \mathbb{R})$, and the real closure K of L with respect to that ordering.

Also solved by T. C. Craven, D. Gay, V. Pambuccian (student, Romania), and the proposer.

The limit $\mathrm{Lim}_n \Pi_i \sin(na_i) = 0$

6336 [1981, 213]. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let $a_1, a_2, \ldots, a_k$ be real numbers, and suppose that

$$\lim_{n\to\infty} \sin(na_1)\sin(na_2) \cdots \sin(na_k) = 0.$$

Prove that at least one of the a_i is an integral multiple of π.

Most of the solvers used some variation of an "equidistribution" argument. From these we present below the solution of Allen Stenger. Others using similar arguments are Ulrich Abel (West Germany), Bela Brindza (Hungary), Robert Breusch, P. L. Chabot, Dinh Thế-Hung & Benny Cheng (students), Michael Golomb, L. Kuipers (Switzerland), O. P. Lossers (Netherlands), Attila

Máté, William A. Newcomb, Victor Pambuccian (Romania), Daniel A. Rawsthorne, Edward Schmeichel, Rodica Simion & Frank W. Schmidt, Abraham Smuckler (Israel), Jospeh Szűcs, and the University of South Alabama Problem Group.

A second type of argument, related to equidistribution but in a setting of dynamical systems, was used by Jonathan J. Ashley, Joe Auslander & Ken Berg, Adam Fieldsteel, and Jose Ramirez-Labrador (Spain).

I. *Solution by Allen Stenger, Hughes Aircraft Company.* Write b_r for a_r/π, and assume that no b_r is an integer. We shall show that the product is bounded away from zero for a positive fraction of the n, and so cannot approach zero in the limit.

Consider first the rational b_r (if any). The portion of the product over these is a periodic function of n, and is nonzero for $n = 1$, so it is bounded away from zero for some positive fraction of the n. We will show that the other part of the product (if any) is also bounded away from zero, for a positive fraction of these n.

Consider now the irrational b_r (if any). Since nb_r is uniformly distributed modulo 1, $\sin \pi n b_r$ is bounded away from zero for a positive fraction of the n; and in fact the fraction can be made as large as desired by making the bound small enough. Therefore the product of $\sin \pi n b_r$ over all irrational b_r is also bounded away from zero for an arbitrarily large fraction of the n.

In particular the fraction can be made so large that there is a positive overlap with the n for which the other part of the product is bounded away from zero.

II. *Solution by R. W. K. Odoni, University of Exeter, England.* The given condition is equivalent to $u_n \to 0$, where $u_n = \prod_{j=1}^{k}(1 - e^{2ina_j})$. It will suffice to prove that $u_n = 0$ for all $n \geqslant 0$. We put $f(z) = \sum_{n \geqslant 0} u_n z^n$; since u_n is bounded, this series certainly defines an analytic function in $|z| < 1$. It will be enough to prove that $f(z) \equiv 0$. We have

$$u_n = \sum_{\varepsilon} \operatorname{sgn}(\varepsilon) e^{2in\sum_j \varepsilon_j a_j},$$

where ε runs over all k-tuples $(\varepsilon_1, \ldots, \varepsilon_k)$ of 0's and 1's, and $\operatorname{sgn}(\varepsilon) = (-1)^{N(\varepsilon)}$ where $N(\varepsilon)$ is the number of 1's in ε. Hence

$$f(z) = \sum_{\varepsilon} \operatorname{sgn}(\varepsilon)\left(1 - z \exp\left(2i \sum_j \varepsilon_j a_j\right)\right)^{-1} \quad [\text{certainly when } |z| < 1],$$

while this formula provides a continuation of f to a function which is meromorphic on the Riemann sphere and has all its poles lying on $|z| = 1$. Suppose we could show that f does not, in fact, have any poles; then f would be constant, while $f(0) = u_0 = 0$, and the proof would then be complete. Choose ϕ real, and let $z = \mathrm{re}^{i\phi}$, $0 < r < 1$. Since $u_n \to 0$ we have $|u_n| < A$ for all $n \geqslant 0$, while $|u_n| < \varepsilon$ for all $n > n_0(\varepsilon)$. It follows that

$$|(1-r)f(z)| \leqslant \sum_{n \leqslant n_0(\varepsilon)} Ar^n(1-r) + \sum_{n > n_0(\varepsilon)} \varepsilon r^n (1-r) \leqslant (1-r)An_0(\varepsilon) + \varepsilon < 2\varepsilon$$

when $1 > r \succ r_0(\varepsilon)$. Hence $\lim_{r \to 1-}(1-r)f(\mathrm{re}^{i\phi}) = 0$ for all choices of ϕ, and f has no poles.

Factorization of n!

6339 [1981, 294]. *Proposed by P. Erdős, Hungarian Academy of Sciences, and J. L. Selfridge, Mathematical Reviews.*

Let $\prod_1^s p_i^{\alpha(i)}$ be the canonical (prime) factorization of $n!$, so that $\pi(n) = s$. Let $f(n) = j$ be the unique index for which

$$\prod_1^{j-1} p_i^{\alpha(i)} < (n!)^{1/2} < \prod_1^{j} p_i^{\alpha(i)}.$$

Prove

(i) $|f(n+1)-f(n)| \leqslant 1$.
(ii) As n increases, each of the relations

$$\prod_1^{j-1} p_i^{\alpha(i)} \gtrless \prod_{j+1}^{s} p_i^{\alpha(i)},$$

has infinitely many solutions.
(iii) $n^{1/2}/p_j$ approaches a fixed limit.
*(iv) For infinitely many n, $f(n+1)=f(n)-1$. For all $m>n$, $f(m) \geqslant f(n)-1$.

Solution by the proposers. Let $a_p(n)$ be the largest exponent e such that p^e divides $n!$, where p is a prime and $p \leqslant n$. Then

$$a_p(n) = \sum_{r=1}^{R} \left[\frac{n}{p^r}\right] \text{ with } R = \left[\frac{\log n}{\log p}\right]. \tag{1}$$

If $p \leqslant n/2$, then $[n/p] \geqslant 2$, and so $a_p(n) \geqslant R+1 \geqslant \log n/\log p$. Hence

$$p^{a_p(n)} \geqslant n \text{ for } p \leqslant \frac{n}{2}. \tag{2}$$

If $p > n/2$, then $p \leqslant n < 2p$ so that $a_p(n)=1$ and hence

$$p^{a_p(n)} = p > \frac{n}{2}.$$

It follows that

$$p^{a_p(n)} > \frac{n}{2} \text{ for } p \leqslant n; \tag{3}$$

and, a fortiori, that

$$p^{a_p(n)} > (n+1)^{1/2} \text{ for } p \leqslant n,\, n \geqslant 5. \tag{4}$$

In view of (1) we have

$$a_p(n) \leqslant \frac{n}{p-1}\left(1-\frac{1}{p^R}\right) \leqslant \frac{n-1}{p-1}; \tag{5}$$

and

$$\begin{aligned} a_p(n) &\geqslant \frac{n}{p-1}\left(1-\frac{1}{p^R}\right) - R \geqslant \frac{n-p}{p-1} - R = \frac{n-1}{p-1} - 1 - R \\ &\geqslant \frac{n-1}{p-1} - 2\frac{\log n}{\log p}. \end{aligned} \tag{6}$$

It follows from (5) and (6) that

$$n^{-2}p^{(n-1)/(p-1)} \leqslant p^{a_p(n)} \leqslant p^{(n-1)/(p-1)},$$

and hence that

$$n^{-2\pi(x)}\exp\left(\sum_{p\leqslant x} \frac{n-1}{p-1}\log p\right) \leqslant F_n(x) \leqslant \exp\left(\sum_{p\leqslant x}\frac{n-1}{p-1}\log p\right) \tag{7}$$

where

$$F_n(x) = \prod_{p\leqslant x} p^{a_p(n)},\, x \leqslant n. \tag{8}$$

Now write $\alpha_i(n) = a_{p_i}(n)$.

Proof of (i). Let $n \geqslant 4$. By (4) and the definition of j, we have

$$\prod_{i=1}^{j+1} p_i^{\alpha_i(n+1)} \geqslant \prod_{i=1}^{j+1} p_i^{\alpha_i(n)} = p_{j+1}^{\alpha_{j+1}(n)} \prod_{i=1}^{j} p_i^{\alpha_i(n)}$$
$$\geqslant (n!)^{1/2}(n+1)^{1/2} = ((n+1)!)^{1/2},$$

whence

$$f(n+1) \leqslant j+1 = f(n)+1.$$

Also, since $(n+1)! = (n+1)\prod_{i=1}^{s} p_i^{\alpha_i(n)}$, we have

$$\prod_{i=1}^{j-2} p_i^{\alpha_i(n+1)} \leqslant (n+1)\prod_{i=1}^{j-2} p_i^{\alpha_i(n)} \leqslant \frac{(n+1)(n!)^{1/2}}{p_{j-1}^{\alpha_{j-1}(n)}} \leqslant ((n+1)!)^{1/2},$$

whence $f(n+1) - 1 \geqslant j - 2$, i.e.,

$$f(n+1) \geqslant f(n) - 1.$$

Thus $|f(n+1) - f(n)| \leqslant 1$ when $n \geqslant 4$. Direct computation shows that the inequality also holds when $n \leqslant 3$.

Proof of (iii). Let $x = p_j$. It follows from (7) and the classical estimate

$$\sum_{p \leqslant x} \frac{\log p}{p-1} = \log x + c + o(1)$$

that

$$\log F_n(x) = (n-1)\sum_{p \leqslant x} \log p + O(\pi(x)\log n)$$
$$= n\log x + cn + o(n) + O(x). \tag{9}$$

Further, since $F_n(x-1) < (n!)^{1/2} \leqslant F_n(x)$, we have

$$\log F_n(x) \sim \tfrac{1}{2}\log n! = \tfrac{1}{2}(n\log n - n) + O(\log n). \tag{10}$$

It follows from (9) and (10) that

$$\log x = \tfrac{1}{2}\log n - \tfrac{1}{2} - c + o(1) + O(x/n) \tag{11}$$

and hence, since $x \leqslant n$, that $x = O(n^{1/2})$. Thus the $O(x/n)$ term in (11) can be absorbed into the $o(1)$ term, and (iii) is an immediate consequence.

Proof of (ii). Let n be such that

$$f(n+1) = f(n) + 1 \tag{12}$$

and

$$p_j^{\alpha_j(n+1)} \geqslant n+1. \tag{13}$$

By (i) and (iii), (12) holds for infinitely many n; and (13) holds for all n sufficiently large by (2) with $n+1$ in place of n. Then

$$\prod_{i=1}^{j-1} p_i^{\alpha_i(n)} < (n!)^{1/2} < \prod_{i=1}^{j} p_i^{\alpha_i(n)}$$

and

$$\prod_{i=1}^{j} p_i^{\alpha_i(n+1)} < ((n+1)!)^{1/2} < \prod_{i=1}^{j+1} p_i^{\alpha_i(n+1)}.$$

Consequently

$$\prod_{j+1}^{s} p_i^{\alpha_i(n)} = \frac{n!}{\prod_{i=1}^{j} p_i^{\alpha_i(n)}} > \frac{n!}{((n+1)!)^{1/2}} = \left(\frac{n!}{n+1}\right)^{1/2}; \tag{14}$$

and, since $\alpha_i(n+1) \geqslant \alpha_i(n)$,

$$\prod_{i=1}^{j-1} p_i^{\alpha_i(n)} \leqslant \frac{\prod_{i=1}^{j} p_i^{\alpha_i(n+1)}}{p_j^{\alpha_j(n+1)}} < \frac{((n+1)!)^{1/2}}{p_j^{\alpha_j(n+1)}} < \left(\frac{n!}{n+1}\right)^{1/2} \tag{15}$$

by (13). The first part of (ii) follows from (14) and (15). To prove the second part we note that when (12) holds

$$\prod_{i=1}^{j} p_i^{\alpha_i(n+1)} \geqslant \prod_{i=1}^{j} p_i^{\alpha_i(n)} > (n!)^{1/2} = \frac{n!}{(n!)^{1/2}} > \frac{n!}{\prod_{i=1}^{j} p_i^{\alpha_i(n+1)}}$$

$$= \frac{1}{n+1} \prod_{i=j+1}^{s'} p_i^{\alpha_i(n+1)} \geqslant \prod_{i=j+2}^{s'} p_i^{\alpha_i(n+1)}$$

by (13), $p_{s'}$ being the largest prime factor of $(n+1)!$.

Proof of part (iv). We shall prove that $f(m) \geqslant f(n) - 1 = j - 1$ for $m > n > n_0$. This amounts to proving

$$\prod_{i=1}^{j-2} p_i^{\alpha_i(m)} < (m!)^{1/2} \quad \text{for } m > n > n_0. \tag{16}$$

We shall in fact prove the stronger inequality

$$\prod_{i=1}^{j-1} p_i^{(m-1)/(p_i-1)} < m! \quad \text{for } m > n > n_0. \tag{17}$$

That (17) implies (16) follows from (5). Let $m > n$. By (6) we have

$$p_i^{\alpha_i(n)} \geqslant n^{-2} p_i^{(n-1)/(p_i-1)}$$

and hence

$$\prod_{i=1}^{j-2} p_i^{\alpha_i(n)} \geqslant n^{2(2-j)} \prod_{i=1}^{j-2} p_i^{(n-1)/(p_i-1)}.$$

Consequently

$$\prod_{i=1}^{j-2} p_i^{(m-1)/(p_i-1)} = \left(\prod_{i=1}^{j-2} p_i^{(n-1)/(p_i-1)}\right)^{(m-1)/(n-1)}$$

$$\leqslant \left(n^{2(j-2)} \prod_{i=1}^{j-2} p_i^{\alpha_i(n)}\right)^{(m-1)/(n-1)} \tag{18}$$

Further

$$\prod_{i=1}^{j-2} p_i^{\alpha_i(n)} = \frac{\prod_{i=1}^{j-1} p_i^{\alpha_i(n)}}{p_{j-1}^{\alpha_{j-1}(n)}} < \frac{(n!)^{1/2}}{p_{j-1}^{\alpha_{j-1}(n)}}. \tag{19}$$

It follows from (18) and (19) that

$$\prod_{i=1}^{j-2} p_i^{(m-1)/(p_i-1)} \leqslant \left(\frac{n^{2(j-2)}(n!)^{1/2}}{p_{j-1}^{\alpha_{j-1}(n)}} \right)^{(m-1)/(n-1)}$$

$$< (m!)^{1/2} \left(\frac{n^{2(j-2)}}{p_{j-1}^{\alpha_{j-1}(n)}} \right)^{(m-1)/(n-1)} \tag{20}$$

since $(n!)^{1/(n-1)} < (m!)^{1/(m-1)}$. In order to establish (17) it suffices, by (20), to prove that

$$n^{2(j-2)} < p_{j-1}^{\alpha_{j-1}(n)} \quad \text{for } n \geqslant n_0. \tag{21}$$

Since $\alpha_{j-1} \geqslant \dfrac{n}{p_{j-1}} - 1$, (21) will follow from

$$n^{2(j-1)} < p_{j-1}^{n/p_{j-1}-1} \quad \text{for } n > n_0. \tag{22}$$

Now, by (iii), $p_{j-1}^3 > (p_j/2)^3 > n$ for n sufficiently large. Thus in order to establish (22) it suffices to prove

$$(6(j-2)+1)p_{j-1} < n \quad \text{for } n > n_0. \tag{23}$$

This follows from (iii) which implies $p_j = O(n^{1/2})$ and that

$$j = O(n^{1/2}/\log n).$$

Apart from removing the restriction $n \geqslant n_0$ in the second part of (iv), the only outstanding result that remains to be established is the first part of (iv). In view of (i) this would be proved if it could be shown that $f(n+1) < f(n)$ for infinitely many n.

Groups with 6 Subgroups

6345 [1981, 353]. *Proposed by Gary L. Walls, University of Southern Mississippi.* Which groups have exactly 6 subgroups?

Solution by Paul L. Chabot, California State University at Los Angeles. The groups which have exactly 6 subgroups are S_3, Q_8, $Z_3 \times Z_3$, Z_{p^5} and Z_{p^2q} where p and q are distinct primes. This is a simple consequence of the following

PROPOSITION. *Let G be a group with at most 6 subgroups. Then G is isomorphic to one of S_3, Q_8, $Z_2 \times Z_2$, $Z_3 \times Z_3$ and Z_n where $n = p^\alpha$ or $n = p^\beta q$, p and q distinct primes, $\alpha \leqslant 5$, $\beta \leqslant 2$.*

Proof. It is clear that G must be finite.

Case 1. Suppose G contains a nonnormal subgroup $\langle x \rangle$ of prime order p. Let $N = N(\langle x \rangle)$ and $g \in G \setminus N$. As $|G : N|$ equals the number of conjugates of $\langle x \rangle$ and there are also the subgroups $\langle 1 \rangle$, G and $\langle g \rangle$, we see that $|G : N| \leqslant 3$. We cannot have $|G : N| = 2$, for then $N \lhd G$ and N would contain a conjugate of $\langle x \rangle$, say $\langle y \rangle$, and hence also $\langle x \rangle \times \langle y \rangle$; and this would entail more than 6 subgroups. Thus $|G : N| = 3$ which implies $N = \langle x \rangle$ and so $p \neq 3$ (as G is not abelian). Hence, we may assume that g is an element of order 3. The 6 subgroups are now present and this forces $\langle g \rangle$ to be normal in G. But as $x \notin C(g)$, we see that $p = 2$ $(= |\mathrm{aut}\langle g \rangle|)$ and $|G| = 6$. Thus $G \approx S_3$.

Case 2. Suppose G is nonabelian and each subgroup $\langle x \rangle$ of prime order is normal in G. Let p be the smallest prime divisor of $|G|$ and let $x \in G$ with $|x| = p$. Since $\langle x \rangle$ is normal in G, we see that $x \in Z(G)$. The quotient $G/\langle x \rangle$ has at most 5 subgroups and is noncyclic since G is nonabelian. By induction, $G/\langle x \rangle \approx Z_2 \times Z_2$. Thus, $p = 2$, $|G| = 8$ and $G \approx Q_8$.

Case 3. Suppose G is abelian. If G is cyclic, then it is easy to check that $G \approx Z_n$ where n is as above. Let G be noncyclic. Then there is a prime p and a subgroup $H \approx Z_p \times Z_p$. Thus, H has $p + 3$ subgroups. We must have $G = H$; otherwise, there would be an element $g \in G \setminus H$ giving at least 7 subgroups, namely, G, $\langle g \rangle$ and the $p + 3$ from H. Also, clearly, either $p = 2$ or 3.

Also solved by Floyd Barger, Leo Comerford, Lorraine L. Foster, Robert Gilmer, Jerrold W. Grossman & Stuart S.-S. Wang, Jay Hook, Michael Josephy (Costa Rica), O. P. Lossers (Netherlands), William Myers, Barbara L. Osofsky, Victor Pambuccian (Romania), Joseph J. Rotman, Dinh Thě Hũng & Benny Cheng, Thomas R. Wolfe, and the proposer.

Joseph J. Rotman and Thomas R. Wolfe point out that the problem is solved in a paper by G. A. Miller, *Groups having a small number of subgroups*, Proc. Nat. Acad. Sci. U.S.A., 25 (1939) 367–371.

Polynomials over GF2

6346 [1981, 353]. *Proposed by M. Slater, University of Bristol, England.*

Sets A_n, B_n $(n = 1,2,\dots)$ of integers are defined recursively as follows:

$$A_1 = \varnothing; \qquad B_1 = \{0\};$$

$$A_{n+1} = \{x + 1 : x \in B_n\}; \qquad B_{n+1} = (A_n \cup B_n) - (A_n \cap B_n).$$

Prove that $B_n = \{0\}$ if and only if n is a power of 2.

Solution by J. G. Mauldon, Amherst College. For $n = 1,2,3,\dots,$ define polynomials $a_n(x)$, $b_n(x)$ in GF2$[x]$ by

$$a_n(x) = \sum_{k \in A_n} t^k \quad \text{and} \quad b_n(x) = \sum_{k \in B_n} t^k \text{ where } t = x(x+1).$$

Then clearly

$$a_n(x) = tb_{n-1}(x),\, b_{n+1}(x) = b_n(x) + a_n(x) \text{ (addition mod 2)},$$

and so by induction

$$a_n(x) = x^n(x+1) + x(x+1)^n \quad \text{and} \quad b_n(x) = x^n + (x+1)^n.$$

Thus $B_n = \{0\}$ if and only if $\binom{n}{r}$ is even for $0 < r < n$, and it is well known and easily proved that this condition holds if and only if n is a power of 2.

Also solved by J. C. Binz (Switzerland), Jozef Bobok (Czechoslovakia), Aage Bondesen (Denmark), Robert Breusch, David G. Cantor, Thomas H. Foregger, Richard Z. Goldstein, Michael Josephy (Costa Rica), William Lambert M. (Costa Rica), Stephen Locke & Bruce Richter (Canada), O. P. Lossers (Netherlands), I. G. Macdonald (England), Jon Merzel, R. W. K. Odoni (England), Vladimír Olejček (Czechoslovakia), Frank Schmidt & Rodica Simion, Bruno O. Shubert, Dinh Thê Hùng and Benny Cheng, Ernst W. Trost (Switzerland), Charles Vanden Eynden, L. Van Hamme (Belgium), Pei Yuan Wu (Taiwan), and the proposer.

MISCELLANEA

88. Easy, simple things are often more valuable than complicated results obtained by hard work.

Hermite to Stieltjes, Feb. 25, 1892.

REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Great Moments in Mathematics (Before 1650). By Howard Eves. The Mathematical Association of America, Washington, D.C., 1980. xiv + 270 pp. List: $22.00; MAA Members: $17.00.

PHILLIP S. JONES
Department of Mathematics, University of Michigan, Ann Arbor, MI 48104

World wide there is a resurgence of interest in the history of mathematics. However, the depth of the interest, the extent of the activity, and the level of mathematics, exposition, and historical research vary considerably. In England the Open University includes courses in the history of mathematics in its curriculum, but has drawn on the United States, France, and Germany for materials. Both R. L. Wilder and the late Carl Boyer appeared on its television broadcasts and Wilder's *Evolution of Mathematical Concepts* was published in a special paperback edition in the Open University textbook series [**8**]. They also published commissioned translations of French [**1**] and German [**7**] books.

In this country, where substantial mathematical research activity is still under one hundred years old, mathematicians are just beginning to realize that the history of mathematics, especially modern, advanced mathematics is interesting, important, and should not be lost. Our Bicentennial Celebration stimulated both special conferences and historical sessions at regular meetings. Earlier American historians of mathematics (David Eugene Smith, Florian Cajori, Louis C. Karpinski, R. C. Archibald) tended to write purely factual-chronological accounts and to concentrate their research on elementary topics. More recently, the range of both American historical exposition and research has been extended, but there are gaps in their coverage of both historical theories and mathematical topics.

Before evaluating the importance of new historical writing, one must analyze the needs and functions of the history of mathematics. These vary with the consuming group. There are three major groups: (1) the professional historians of mathematics and science, (2) mathematicians who have become interested in their "roots" or in the relationships between specialities, (3) teachers of mathematics and of the history of mathematics. Of course, these categories are neither complete nor mutually exclusive. A person may fit more than one or may shift his category from time to time. Further, a fourth group could be added, that portion of the general public which pursues scientific and historical interests. Recall the periodic appearance of mathematical books on "best seller" lists, Lancelot Hogben's *Mathematics for the Million*, E. T. Bell's *Men of Mathematics*, and, most recently, Douglas Hofstadter's *Gödel*, *Escher*, *Bach*.

Although courses and seminars in history for graduate students in mathematics exist abroad, in the United States the clientele for the history of mathematics has largely been teachers, chiefly undergraduate, preservice teachers. This affects the level and emphasis of the course, but should not significantly affect its goals. The major goal for the history of mathematics course should be to help the student to better his understanding of the mathematics of the *present* as a preparation for better understanding in doing, using, and teaching mathematics in the future. The important understandings that exposure to history can convey better than is done in most mathematics courses are: (a) a perception of the current views of the *foundations* and *nature* of mathematics as a product of changes from the premathematical stage of "subconscious" use of numbers and shapes, to modern emphasis on axiomatics, logic, and structure; (b) an insight into the varied *motivations* and sources for new mathematical ideas; (c) a view of different *methods* used or steps taken in doing mathematics; (d) the many *interrelationships* within mathematics and with the physical world, philosophy, logic, even art and beauty, which explain the different *meanings* which

mathematics has for different people.

Of course none of these somewhat esoteric goals or "understandings" can be approached without a basic knowledge of both *mathematics* and the facts of its *history*. This factual content is a sine qua non of historical writing.

The book which furnishes the occasion for this essay-review is aimed at our category 3 consumer, the teacher-student or student-teacher group. For this group the history of mathematics has two additional attitudinal values, the intrinsic interest of stories about intellectual adventures in earlier years and anecdotes which reveal the human nature and perhaps foibles of persons who otherwise seem unreal and far removed. We need to combat, especially in the schools, the view that mathematics is an esoteric activity entirely revealed to and used by other-worldish people, beyond the ken of ordinary people. However, the history of mathematics should not be equated solely to facts and anecdotes. Eves' book is full of facts and anecdotes as are his well-known *Introduction to the History of Mathematics* [**2**] and his *Mathematical Circles* [**3**][**4**]. In fact this volume in a sense antedates all of them. It is a condensation of twenty lectures including twenty-three "great moments" selected from thirty semiweekly lectures in a one-semester "appreciation" course for college students with an acquaintance with high school mathematics. There is a set of problems, many quite challenging, following each lecture. Students who elected to attend a weekly problem session based on them could earn a third hour of credit for the course.

The book could also be used for optional enrichment reading-problem assignments (there are hints and answers for many of the problems) in elementary or survey courses or freshman seminars. It is not a text for a history course but might be used for extra reading-report-problem assignments in methods courses containing a history unit.

The book is not solely factual and anecdotal. The "great moments" are well selected. They are interesting and they do, implicitly, furnish illustrations of the "understandings" which we listed earlier. However, unless the illustrations are pointed out, amplified and discussed as such, the generalizations will be perceived only by unusual students. For example, *motivations* for doing mathematics are implicit in such statements as "all significant and lasting mathematics is born of and nurtured by the real world" (p. 194) in connection with the work of Galileo; but it should be contrasted with, "One never knows when a piece of pure mathematics may receive an unexpected application" (p. 198) in connection with Kepler's use of properties of conics developed by the Greeks, and with "growth in mathematics is stimulated by the presence of challenging problems" (p. 62) noted in connection with the three famous Greek problems.

Similarly, Eves' examples of extensions of the Pythagorean theorem suggest that *generalization* is typical of mathematical growth, especially in modern times. However, explicit discussion and additional illustrations are needed to make its nature clear, its importance recognized, and its connection with *abstraction* understood.

American historians of mathematics have paid little attention to theories of discovery and growth such as contained in Kuhn's *The Structure of Scientific Revolutions* [**5**], Lakatos' *Proofs and Refutations* [**6**], and Wilder's views of cultural stress and mathematical evolution [**8**]. However, Eves was on the verge of such ideas in Lecture 5, "The First Crisis" (the irrationality of $\sqrt{2}$), Lecture 6, "Resolution of the First Crisis" (Eudoxus' definition of equal ratios), and Lecture 7, "First Steps in Organizing Mathematics" where he does mention two theories as to the origin of the axiomatic method, *evolutionary* and *revolutionary*.

These comments are not criticisms of this book. Given its purposes, projected audience and space limitation, it could hardly be better. It is clear that in its sequel Eves will follow up, in time, some of the themes begun in this volume. However, for the history of mathematics to function as a significant aid to teaching, as the International Study Group on the Relations between the History and Pedagogy of Mathematics [**9**] believes it can, we need more time for teaching it, more discussion of such issues as are noted here, together with more supporting examples of them and more usable, translated and annotated source materials to support the current texts.

References

1. P. Dedron and J. Itard, Mathematics and Mathematicians, Transworld Publishers, London, 1974, vol. 1, 323 pp., vol. 2, 222 pp. First published as Mathématiques et Mathématiciens, Editions Magnard, Paris, 1959.

2. Howard Eves, An Introduction to the History of Mathematics, 4th ed., Holt, Rinehart, and Winston, New York, 1976, xv + 588 pp.

3. ______, In Mathematical Circles, Prindle, Weber & Schmidt, Boston, 1969, vol. 1, xvii + 136 pp., vol. 2, xi + 145 pp.

4. ______, Mathematical Circles Revisited, Prindle, Weber & Schmidt, Boston, 1971, xxii + 186.

5. Thomas S. Kuhn, The Structure of Scientific Revolutions, 2nd ed., University of Chicago Press, Chicago, 1970, xii + 210.

6. Imre Lakatos, Proofs and Refutations, Cambridge University Press, London, 1976, xii + 174 pp.

7. Walter Popp, History of Mathematics: Topics for Schools, Milton Keynes, England: The Open University Press, 1978, x + 150 pp. Originally published as Geschichte der Mathematik im Unterricht, Bayerischer Schulbuch Verlag, Munich, 1968.

8. Raymond L. Wilder, Evolution of Mathematical Concepts, Wiley, New York 1968, xiii + 224, Open University edition, 1979, 240 pp.

9. This study group currently chaired by Bruce E. Meserve, University of Vermont, and Roland Stowasser, Technische Universität Berlin, is affiliated with the International Congress on Mathematics Education (ICME).

Deterministic Mathematical Models in Population Ecology. By H. I. Freedman. Marcel Dekker, New York, 1980.

MAYNARD THOMPSON
Department of Mathematics, Indiana University, Bloomington, IN 47405

The study of mathematical systems which have been proposed as models for biological phenomena is a flourishing research activity. Even a brief look through the Biology and Behavioral Science section of *Mathematical Reviews* verifies the level of activity and the diversity of both the biological problems and the mathematical methods. Among the areas which have been and continue to be especially vital are those related to the study of populations. The scientific interest in this work arises partially from the variety of applications (economics, demography, genetics, ecology, resource management, etc.) and partially from the insight it provides into the underlying scientific principles. From the mathematical side, interesting problems in differential equations, topological dynamics, discrete dynamical systems, stochastic processes and related fields have arisen in the study of models for populations.

In constructing a model for a community (a set of subpopulations) it is natural to begin by looking at the internal dynamics of each subpopulation. There are choices to be made: a deterministic or stochastic model, spatial and temporal variation or temporal variation only, continuous or discrete time (space), functional form of vital dynamics, etc. Although single species population models have the longest history, new problems and new results continue to emerge. Thus, even for deterministic models there are many alternatives to the simple Malthusian ($\dot{x} = bx$) and Verhulst ($\dot{x} = bx(1 - x/K)$) differential equations. For instance, a simple nonlinear discrete time model for a single species population with nonoverlapping generations leads to the difference equation $x_{t+1} = bx_t(1 - x_t)$. This equation, and similar ones, describes systems whose dynamical behavior is remarkably complex: the trajectory $\{x_t\}$ of such a system can exhibit stable fixed points, stable cyclic oscillations, and completely aperiodic behavior. Robert May discussed such systems from a scientific point of view in a series of articles in *Science* and *Nature* beginning in 1974, and Li and Yorke pursued the mathematics in a MONTHLY article in 1975. In another direction, one can consider models incorporating age dependent vital dynamics and the

resulting linear or nonlinear systems. Again, much has been learned about such systems in the 1970's.

Once one knows something of the behavior of a single population, it is natural to consider the (biologically more significant) situation in which two or more populations interact. From a mathematical point of view, the case of two interacting populations is especially tractable. There is also some biological justification for paying special attention to communities of two populations. In the 1920's Lotka and Volterra independently proposed models for interacting populations of the predator-prey type:

$$\dot{x} = x(\alpha - \beta y), \qquad \dot{y} = y(-\gamma + \delta x), \qquad \alpha, \beta, \gamma, \delta > 0.$$

The study of these models led to the consideration of a variety of other types of interactions, functional forms of interaction terms, numbers of interacting populations, etc.

Freedman's book surveys a variety of deterministic models for interacting populations. The focus is on two-population, continuous time models; that is, systems of two ordinary differential equations. In addition to Lotka-Volterra predator-prey models, models for competition, cooperation, and Gauss's generalization ($\dot{x} = x - yp(x)$, $\dot{y} = y(-\gamma + cp(x))$) are also discussed. More general situations, models of Kolmogorov type: $\dot{x} = xf(x, y)$, $\dot{y} = yg(x, y)$, are introduced and their properties are described, but details of proofs are omitted. Many of the discussions are concerned with the existence and stability of equilibria, and the phase plane analysis of the systems. There are also multispecies models with more than two subpopulations: three-species food chains, one predator-two prey, and two predator-one prey communities are examples. Some of the discussions are complete, Lotka-Volterra systems for instance, while others outline the model and its analysis, but refer to the literature for details. The documentation is especially thorough, and there are helpful Notes on the Literature sections in each chapter. The bibliography of nearly 1000 items is a major contribution in itself. A number of open questions are noted.

This book presents the mathematics associated with certain population models and it provides a tightly focused introduction to the subject. As a result of the emphasis on models leading to systems of ordinary differential equations, discrete time models receive little attention, and models with age dependent vital rates are omitted completely. It is written in a style which most mathematicians will find comfortable, though not every definition, theorem, and proof is labeled as such. A reader might find a book which discusses the biological setting, e.g., Robert May, *Stability and Complexity in Model Ecosystems* (Princeton, 1974) a useful companion volume.

ACKNOWLEDGMENT

The editors thank the following persons who have refereed manuscripts between mid-1981 and mid-1982: J. Adney, G. L. Alexanderson, R. Allen, R. Anderson, S. S. Anderson, G. E. Andrews, T. S. Angell, K. I. Appel, T. Armstrong, E. A. Azoff, A. Back, P. Bateman, J. Baxley, L. Baxter, E. F. Beckenbach, G. Bennett, M. K. Bennett, L. D. Berkovitz, B. Berndt, R. A. Bieberich, D. Bindschadler, R. Borden, J. M. Borwein, P. B. Borwein, T. A. Botts, R. Bradley, D. S. Bridges, A. Brown, P. J. Browne, E. Buchman, R. C. Buck, B. Carlson, J. Cederverg, R. A. Chaffer, A. Charnow, W. Cheney, S. -N. Chow, W. W. Comfort, C. Cowen, H. S. M. Coxeter, H. H. Crapo, F. H. Croom, T. Cusick, J. Dadok, J. W. Dauben, J. Daughtry, E. Davis, M. D. Davis, P. J. Davis, R. DeBruin, C. R. Deprima, R. L. Devaney, C. DeVito, S. Z. Ditor, L. E. Dubins, R. Dudley, U. Dudley, J. R. Durbin, J. Dyer-Bennett, H. M. Edgar, H. Edwards, P. Erdős, L. Erlebach, M. Evens, H. Eves, E. R. Fadell, K. Fan, H. E. Fettis, J. Fife, W. Firey, H. Flanders, J. Foster, L. Foster, R. B. Frary, D. Gale, J. Gallian, J. Galovich, S. Galovich, J. Garfunkel, R. Geroch, L. Gillman, S. Goldberg, J. Goldfeather, S. W. Golomb, B. Gordon, H. W. Gould, J.

Grabiner, S. Grabiner, R. L. Graham, P. Greene, R. Greenwood, H. Greshenson, D. S. Griffeath, C. M. Grinstead, B. Grünbaum, H. Guggenheimer, A. P. Guinand, H. Halberstam, R. W. Hamming, K. Harrow, J. Hartman, W. S. Hatcher, M. Hausner, D. Hayes, J. R. Hedstrom, M. Henriksen, C. W. Henson, I. N. Hernstein, J. V. Herod, G. A. Heuer, E. Hewitt, P. Hilton, P. G. Hinman, K. M. Hoffman, S. -P. Hodgson, D. Hofstadter, R. B. Holmes, R. Honsberger, R. E. Howe, P. Humke, J. Hutchinson, E. Ingraham, J. Isbell, M. E. -H. Ismail, M. Q. Jacobs, J. E. Jamison, C. G. Jockush, Jr., C. R. Johnson, P. E. Johnson, R. Kakigi, W. Kaufmann-Buehler, N. D. Kazarinoff, H. J. Keisler, E. Keller, L. M. Kelly, P. J. Kelly, J. G. Kemeny, R. J. Kinney, R. B. Kirchner, R. B. Kirk, M. Klamkin, R. S. Kleber, V. L. Klee, Jr., N. I. Koblitz, J. B. E. Konhauser, W. E. Lakey, T. Lenker, D. Lin, C. Lo, W. Loud, S. Lubkin, W. A. J. Luxemburg, S. Mac Lane, D. P. Maki, J. Mallet-Paret, B. Manvel, M. Marden, J. Marik, J. Marsden, J. M. Masley, J. Masterson, R. McConnell, J. McIntosh, R. M. McKenzie, J. E. McLaughlin, P. McMullen, R. Merriss, E. A. Michael, R. S. Millman, G. Mills, R. Moeckel, S. Montgomery, G. Morrow, Y. N. Moschovakis, W. O. J. Moser, P. S. Mostert, D. Mugler, R. C. Mullin, S. B. Nadler, D. Nance, A. Nijenhuis, T. Nishiura, I. Niven, E. Nssah, E. Nummela, A. M. Odlyzko, I. Olkin, A. Ostebee, J. C. Owings, Jr., E. W. Packel, G. D. Parker, B. Parlett, C. J. Parry, J. Pedersen, M. J. Pelling, G. Phillips, J. M. Plotkin, A. van der Poorten, M. Pour-El, C. Prather, W. M. Priestley, M. Protter, S. E. Puckette, M. O. Reade, C. M. Reis, J. Reneke, H. J. Reyser, J. W. Rice, J. I. Richards, A. W. Roberts, R. M. Robinson, R. T. Rockafellar, D. A. Rose, D. P. Roselle, M. Rosen, G. Rosenstein, K. Ross, J. J. Rotman, W. Rudin, R. Ryden, H. J. Ryser, D. G. Saari, T. L. Saaty, B. Sagan, J. J. Schäffer, R. Schneider, M. Schreiber, A. Schurle, J. L. Selfridge, J. Shallit, G. J. Sherman, A. Shields, J. R. Shoenfield, R. A. Shore, M. Shub, S. G. Simpson, D. Singmaster, D. D. Smith, K. Smith, K. T. Smith, M. K. Smith, M. Sobel, L. Sons, J. Spanier, R. St. Andre, I. Stakgold, H. Stark, J. M. Steele, S. K. Stein, A. Sterrett, K. B. Stolarsky, R. Sverdlove, O. Taussky-Todd, M. Thompson, R. M. Thrall, M. Tomber, L. Tomm, P. C. Trombi, A. Tucker, N. D. Turner, C. L. Vanden Eynden, E. van Douwen, T. P. Vaughan, S. S. Wagstaff, Jr., M. Wand, C. Weil, B. M. Weinstock, F. Weissler, J. G. Wendel, R. J. Whitley, C. T. Whyburn, A. Wilansky, B. J. Winkel, M. Wonenburger, P. Zorn.

MISCELLANEA

89.

... as mathematics, a human invention
That parallels but never touches reality, gives the astronomer
Metaphors through which he may comprehend
The powers and the flow of things: so the human sense
Of beauty is our metaphor of their excellence, their
divine nature:—like dust in the whirlwind, making
The wild wind visible.

—Robinson Jeffers. Reprinted, by permission, from the poem "The Inhumanist," in *The Double Axe and Other Poems*, Liveright, New York, 1977. (Suggested by M. F. Smiley, who considers this the best published answer to the question, "What is Mathematics?")

REVIEWS

Names of authors are in ordinary type; those of reviewers, in capitals.

FILMS

ERRATA

Vol. 89

Kenschaft, "Black Women in Mathematics in the United States," October, 1981: In attempting to list all American Black women with doctorates in mathematics, the author missed two women. *Annie Marie Garraway* received her Ph.D. from the University of California at Berkeley in 1967 and is now employed by Bell Laboratories in Columbus, Ohio. *Emma Rose Fenceroy* received her degree from the University of Alabama in 1979 and is now head of the Department of Mathematics at Florida A & M University in Tallahassee. Three more Black women have completed their degrees since the article went to press.

U S POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U S C 3685)

1 TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY

A PUBLICATION NO: A M M Y A E

2 DATE OF FILING: Sept 16, 1982

3 FREQUENCY OF ISSUE: Monthly except July and August

A NO OF ISSUES PUBLISHED ANNUALLY: Ten

B ANNUAL SUBSCRIPTION PRICE: Library $50 00; Indiv Member $20 00

4 COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street City County State and ZIP Code) (Not printers)*: 1529 18th Street, NW, Washington, DC 20036

5 COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: 1529 18th Street, NW, Washington, DC 20036

6 FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER EDITOR AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*: The Mathematical Association of America, 1529 18th St ,NW, Washington, DC 20036

EDITOR *(Name and Complete Mailing Address)*: P R Halmos, Indiana University, Dept. of Mathematics, Bloomington, IN 47405

MANAGING EDITOR *(Name and Complete Mailing Address)*: Dr Raoul Hailpern, SUNY at Buffalo, Buffalo, NY 14214

FULL NAME	COMPLETE MAILING ADDRESS
The Mathematical Association of America,Inc	1529 18th St., NW Washington, DC 20036

8 KNOWN BONDHOLDERS MORTGAGEES AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS MORTGAGES OR OTHER SECURITIES *(If there are none so state)*

FULL NAME	COMPLETE MAILING ADDRESS
None	

9 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 411 3, DMM only)*
The purpose function and nonprofit status of this organization and the exempt status for Federal income tax purposes (*Check one*)

[X] (1) HAS NOT CHANGED DURING PRECEDING 12 MONTHS
[ ] (2) HAS CHANGED DURING PRECEDING 12 MONTHS
(If changed publisher must submit explanation of change with this statement)

10 EXTENT AND NATURE OF CIRCULATION	AVERAGE NO COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A TOTAL NO COPIES *(Net Press Run)*	21,181	20,824
B PAID CIRCULATION 1 SALES THROUGH DEALERS AND CARRIERS STREET VENDORS AND COUNTER SALES	none	none
2 MAIL SUBSCRIPTION	18,344	16,220
C TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	18,344	16,220
D FREE DISTRIBUTION BY MAIL CARRIER OR OTHER MEANS SAMPLES COMPLIMENTARY AND OTHER FREE COPIES	77	77
E TOTAL DISTRIBUTION *(Sum of C and D)*	18,421	16,297
F COPIES NOT DISTRIBUTED 1 OFFICE USE LEFT OVER UNACCOUNTED SPOILED AFTER PRINTING	2,760	4,527
2 RETURN FROM NEWS AGENTS	none	none
G TOTAL *(Sum of E F1 and 2 should equal net press run shown in A)*	21,181	20,824

11 I cer[illegible]
me [illegible]

PS Form July 1981 3526

(See instruction on reverse)

(Page 1)

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.